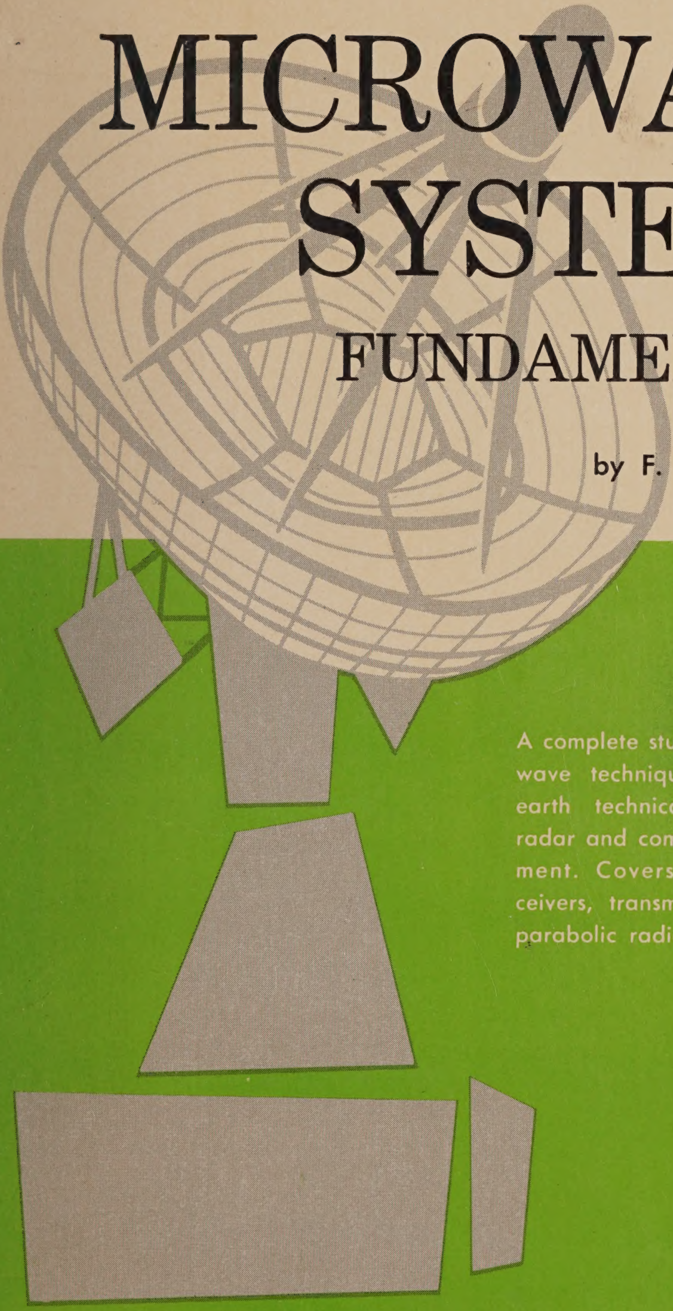
A Howard W. Sams PHOTOFACT PUBLICATION

MICROWAVE SYSTEMS FUNDAMENTALS

by F. Jonathan Mivec



A complete study course in microwave techniques, with down-to-earth technical information on radar and communications equipment. Covers transmitters, receivers, transmission-line systems, parabolic radiators, etc.

$5.95

Cat. No. MSM-1

MICROWAVE SYSTEMS FUNDAMENTALS

by

F. Jonathan Mivec

HOWARD W. SAMS & CO., INC.
THE BOBBS-MERRILL COMPANY, INC.
Indianapolis • New York

FIRST EDITION

FIRST PRINTING—FEBRUARY, 1963

MICROWAVE SYSTEMS FUNDAMENTALS

Copyright © 1963 by Howard W. Sams Inc., Indianapolis 6, Indiana. Printed in the United States of America.

Reproduction or use, without express permission, of editorial or pictorial content, in any manner, is prohibited. No patent liability is assumed with respect to the use of the information contained herein.

Library of Congress Catalog Card Number: 63-12598

Preface

Microwave theory is not new; it has been known by scientists and students of advanced electronics for many years. Certainly, microwave equipment is no stranger to most of us. Airports use microwave radar to keep a visual check on aircraft, telegraph companies use microwaves to handle thousands of cross-country telephone conversations and radio and television programs, law enforcement agencies use radar to monitor auto speeds, and microwaves are used in precise measurements of certain equipment in industry.

In order to fully understand the concepts of microwave theory, it is necessary to have a good background in radio theory. Microwaves are actually not much different from frequencies in the VHF and UHF ranges; however, the associated components are often quite different. From this book, you will learn all the key procedures and practices used in modern microwave equipment. From a fundamental study of microwaves, you will be presented with a complete and detailed break-down of every major component used in microwave communications and radar systems.

This book will take you into what I consider the fascinating world of microwaves—a very new and, as yet, a very enthralling field for the average radio technician to study. By learning the principles presented herein, and by carefully studying the material on microwave equipment, you will have a very sound working knowledge of modern microwave equipment.

I would sincerely like to thank those of my close friends and associates who read portions of this manuscript. I especially want to thank my daughter, who devoted many hours and broken fingernails to typing this manuscript. In addition, I wish to thank Mr. E. H. McMichael of the Department of The Army in Washington, D. C. for granting me permission to use material from Army Technical Manuals in the production of this book.

F. JONATHAN MIVEC

January, 1963

Contents

Chapter 1

Introduction to Microwaves 7

Microwave Aspects — Standard Waveguides — Microwave-Component Dimensions—Bandwidth and Directivity of Microwave Systems—Microwave Antennas

Chapter 2

Behavior of Microwaves 14

Paths of Radio Waves—Curvature of Microwave Paths—Atmospheric Conditions—Importance of Reflected Signal—Transmitting-Antenna Directivity—Horizontally and Vertically Polarized Waves—Effect of Wave Paths on Communications Installations—Effect of Wave Paths on Search-Radar Installations—Line-of-Sight Limitations—Other Limitations of Maximum Range—Maximum Range of Point-To-Point Communications System—Maximum Range of Radar System

Chapter 3

Microwave Transmission Systems 51

Types of Transmission Lines—Electric and Magnetic Fields in Coaxial Line—Waveguides—Group and Phase Velocities—Other Modes of Transmission — Rectangular-Waveguide Modes—Circular-Waveguide Modes—Waveguide Advantages—Tuning Devices and Related Equipment

Chapter 4

Cavity Resonators 101

Reinforcement of Signals—Simple Resonator—Principles of Resonant Circuits — Rectangular Resonators — Cylindrical Resonators—Other Resonators—Cavity-Resonator Tuning—Coupling to Cavity Resonators—Types of Cavity Resonators

Chapter 5

Microwave Transmitter Oscillators 132
Klystrons and Magnetrons — Feedback Oscillators — Relaxation Oscillators — Ringing Oscillators — Conventional-Tube Oscillators—Transformer-Coupled Oscillator—RC Oscillator—Three-Reactance Oscillators—Velocity Modulation—Klystron Oscillators—Cavity Magnetrons

Chapter 6

Microwave Receivers 186
Frequency Conversion—Microwave Frequency Mixers—Local Oscillators — Frequency-Converter Assemblies — Radar Beacons—Search-and-Beacon Frequency - Converter Assembly—Preselection—Intermediate - Frequency Amplification—Detection and Amplification—Automatic Gain and Frequency Control—Special Aspects of Communications Receivers

Chapter 7

Microwave Antennas 221
Transmission Line-to-Load Impedance Matching—Generator-to-Line Impedance Matching — Line-to-Space Impedance Matching—Microwave Antennas—Dissipative Terminations—Radiation Patterns—Radiation From Line Source—Radiation From a Two-Dimensional Source—Radiation From Line Array—Two-Dimensional Array Radiation—Interpretation of Radiation Patterns — Array Method of Antenna Design (Using Reflectors)—Special Reflectors—Reflector Feed Systems—Metal Lenses—Scanning and Stabilization

Appendix I

Mathematical Construction of Profile Diagram 262

Appendix II

Waveguide Mathematics and Equivalent Circuits
of Amplifier and Quartz Crystal 265

Index..282

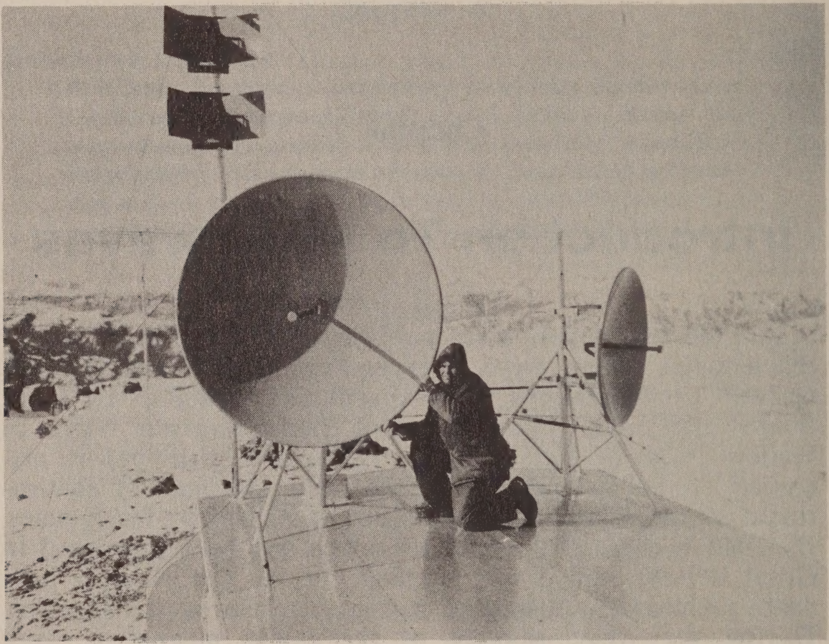
Chapter 1

Introduction To Microwaves

In 1936 two papers were published that dealt with hollow waveguides. These were, "Hyper-Frequency Wave Guides–General Considerations and Experimental Results," by G. C. Southworth; and, "Transmission of Electromagnetic Waves in Hollow Tubes of Metal," by W. L. Barrow. Both papers discussed the hollow metal tubes that take the place of conventiónal transmission lines in operation at microwave frequencies. By 1940, a successful cavity magetron had been developed in Great Britain, and it was possible, for the first time, to generate substantial amounts of power at microwave frequencies. During that year, samples of the cavity magnetron were brought to the United States and research and development work in the microwave field was begun in the Radiation Laboratory at the Massachusetts Institute of Technology.

Much of the early work at the Radiation Laboratory was directed toward the design of airborne microwave radar equipment, since operation at microwave frequencies permits the use of relatively small antenna structures. The term *radar* actually is a colloquialism for microwave equipment, although it refers to but one type. That type is that which is used to *see* electronically, by means of a transmitted radio wave that reflects from the seen object, the reflection being picked up by a receiver and translated into some indication of the range of the object, its bearing, and if necessary, its elevation angle. In the beginning, however, the first practical radar equipment operated in the VHF range, and the radar used in the early days of World War II operated at those frequencies.

Wartime development work in the microwave field was directed primarily toward improvement of radar equipment of various sorts. Since the war, however, various types of microwave communications equipment have come into use. Most of these are microwave relay installations which handle telegraph, telephone, or television signals. Fig. 1-1 shows a telephone relay installation designed for short-distance work, and Fig. 1-2 shows a microwave television relay installation. Operation at microwave frequencies offers two major advantages—the great directivity of microwave antennas permits the use of relatively low-power transmitting equipment, and the great bandwidth available at microwave frequencies affords tremendous traffic-handling capacity. In both respects, microwave equipment is superior to VHF equipment. A single low-power telephone relay installation operating at microwave frequencies, for example, is able to handle more than a thousand simultaneous conversations.



Courtesy Raytheon Mfg. Co.

Fig. 1-1. Microwave telephone relay installation.

Courtesy Raytheon Mfg. Co.

Fig. 1-2. Microwave television relay installation.

MICROWAVE ASPECTS

The difference between microwave theory and ordinary circuit theory may be compared to the difference between DC and AC circuit theory. After studying DC theory, it is necessary to acquire a new point of view in order to understand AC theory, with concepts like phase, frequency, and reactance which bear little apparent relation to DC theory.

Conventional radio circuit theory is based on voltages and currents, but much of microwave theory is based on electromagnetic fields. This actually is not a great change in thinking. To begin with, the concept of electromagnetic fields forms the basis of all antenna theory. Also, once the operation of waveguides and their associated components is grasped, they may be considered from the familiar viewpoint of transmission-line theory. The standard concepts of impedance matching and standing-wave ratio apply to waveguide installations just as they apply to conventional transmission-line systems. Because of this, there is a close similarity between microwave and VHF systems. This becomes apparent as soon as the fundamentals of waveguide theory have been digested. Just as AC theory is a logical extension of DC theory, so microwave practices and techniques follow easily and simply from those of conventional radio.

STANDARD WAVEGUIDES

Fig. 1-3 shows two standard types of waveguides, rectangular and circular. At this point, it is necessary to discard the concept of transmission-system operation in terms of voltages and currents. Instead, anyone understanding electric and magnetic fields will be close to an understanding of the microwaves that travel inside the waveguide. There is simply a change in the terms and the principles used in thinking about transmission systems. Once this change in approach is accepted, a description of waveguides and other components of microwave transmission systems will be grasped easily.

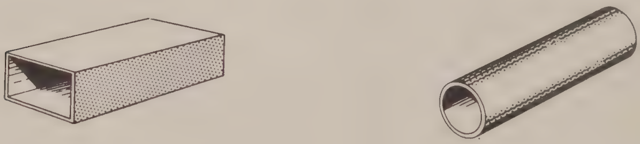

(A) Rectangular. (B) Circular.

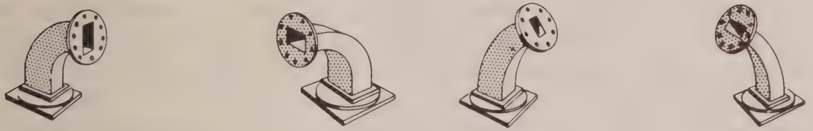
Fig. 1-3. Standard types of waveguides.

MICROWAVE-COMPONENT DIMENSIONS

The dimensions of most microwave components are nearly proportional to the wavelength at which they operate, and therefore the size of the equipment varies with the wavelength of the signals. For obvious reasons, microwave systems are usually designed for operation in one of the three standard bands given in Table 1-1, and components normally are available in sizes adapted to the three wavelengths.

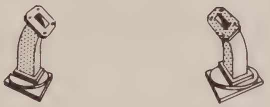
Table 1-1. Microwave frequency versus wavelength.

Bands	Frequencies (mc)	Wavelengths	
		(cm)	(in)
S-band	Near 3,000	10	4
X-band	Near 10,000	3	1.2
K-band	Near 27,000	1.1	.44



(A) S band.

(B) X band.



(C) K band.

Fig. 1-4. Standard sizes of waveguide bends.

The three standard sizes of a bent section of waveguide are shown in Fig. 1-4. The largest section is used for the S-band, the middle size for the X-band, and the smallest size for the K-band. The flange fittings on the end of each section are used for connection to adjacent pieces of waveguide.

Although signals in the S-, X-, and K-bands behave in similar ways, the effects of rain, snow, and fog are more pronounced in the higher-frequency X- and K-bands. For this reason, early-warning radars which must pick up targets at long ranges under all weather conditions are usually operated at S-band frequencies. Most airborne radars operate at X-band frequencies to take advantage of the weight savings afforded by the reduction in size of microwave components. Antiaircraft tracking radars also operate at X-band frequencies, sacrificing long-range, bad-weather performance to obtain the greater tracking resulting from the sharper beam obtainable in the X-band. The K-band is reserved for special applications, such as radars, where light weight and small antenna size are more important than reliable long-range operation. A typical example of K-band equipment is the midget homing radar used in a guided missile.

BANDWIDTH AND DIRECTIVITY OF MICROWAVE SYSTEMS

One of the advantages of operation at microwave frequencies is the great bandwidth of microwave systems or, in other terms, the tremendous amount of information that can be carried by a single microwave system. When an RF carrier frequency (f_c) is amplitude-modulated by an audio signal frequency (f_a), the resulting modulated carrier consists of three distinct RF signals. The first of these is the original carrier. The others, at frequencies f_a above and below the carrier, are called the *upper and lower side bands*, respectively. If the carrier is frequency modulated, there are additional sidebands at frequencies $(f_c + 2f_a)$, $(f_c - 2f_a)$, $(f_c + 3f_a)$, $(f_c - 3f_a)$, and so on. The RF section of any transmitter is, necessarily, tuned to the carrier frequency (f_c) and will pass only signals having frequencies within about 5% of f_c. (The limit is, of course, different with different types of equipment.) In practice, f_a is often limited to values such as .01 f_c. It follows then that higher modulating frequencies can be used when the frequency of the carrier is increased.

Consider, for example, the 96 channels at frequencies of 550 kc, 560 kc, 570 kc, and so on up to 1,500 kc. Operating on these channels, 96 transmitters can carry 96 separate voice programs. It is possible, however, to operate 96 low-power RF amplifiers at these frequencies and modulate each with its own voice program. The outputs of the 96 modulated amplifiers then can be combined (by adding them together) into a single, rather complex, signal containing frequencies that range from slightly below 550 kc to slightly above 1,500 kc. This complex signal then can be used to modulate a microwave transmitter. It is therefore possible to carry all 96 voice programs in the *entire* broadcast band on a single microwave carrier. At the receiving end of the microwave system, routine detection of the incoming signal provides the complex signal (containing the 96 modulated carriers) at a high level with a negligible noise background. After this, 96 simple receiver-type circuits then select the separate modulated carriers, and the outputs of these circuits are detected in the usual way to obtain the 96 voice programs—each on a separate line. Thus, through the use of more refined techniques (basically the same as the technique described above) a single microwave relay installation is enabled to handle over a thousand telephone conversations at the same time.

Another advantage of operation at microwave frequencies is the great directivity that can be obtained with relatively small antenna systems. In general, the sharpness of an antenna pattern depends on the size of the antenna. If two transmitting antennas are operated at the same frequency, the larger one ordinarily will radiate a narrower beam than the smaller one. Similarly, if two receiving antennas are operated at the same frequency, the larger will have a greater directivity than the smaller. It turns out, however, that two antennas operating at different frequencies will have similar radiation or similar reception patterns if their dimensions in wavelengths are similar. As an example, an antenna array 10 feet high and 20 feet wide operating at a wavelength of 1 meter (at a frequency of 300 mc) has essentially the same radiation pattern as an array 20 feet high and 40 feet wide operating at a wavelength of 2 meters (at a frequency of 150 mc). Thus, microwave antennas of small size may achieve the same radiation pattern as the larger ones used in VHF systems.

MICROWAVE ANTENNAS

One type of early-warning radar, which was used extensively before microwave equipment was developed, operated at a frequency near 700 mc and used an antenna array about 6 feet wide by nearly 6 feet high. This array was a heavy and awkward amount of gear to swing around on the top of an antenna tower. The machinery necessary to turn it and the structure required to support it were necessarily large and expensive. With microwave equipment, however, a similar pattern can be obtained at S-band frequencies with an antenna approximately 16 inches wide by 12 inches high, or with a circular structure slightly over 6 inches in diameter. In practice, an S-band antenna is rarely as small as this, but it is obvious that the savings in weight and size may be considerable.

Antenna sizes for airborne radars are generally about 30 inches in diameter for the S-band, and anywhere from 8 to 30 inches for the X-band. A present-day, early-warning radar operating in the S-band uses an antenna about 3 by 10 feet. This is approximately one-quarter the size of the VHF radar mentioned earlier, but the efficiency of the S-band antenna is so increased that there is no comparison between its performance and that of its VHF forerunner.

Chapter 2

Behavior Of Microwaves

It usually is possible to obtain reliable operation over distances up to 100 miles with radio systems operating at frequencies below 30 mc, but this is not true with systems operating at microwave frequencies. Even with a powerful transmitter and a good antenna system, one cannot be sure of delivering a usable signal at a specified point. In some cases, for example, a change of 50 feet in the height of the transmitting antenna may provide excellent operation where *no operation* was possible before. In other cases, communication between two points may be possible only under abnormal weather conditions. Such behavior results from variation in the *propagation characteristics* of the path joining the transmitter to the desired receiving point.

Given two points, the microwave propagation characteristics of the path between them—that is, the way in which microwave energy travels from one point to another—depend on such factors as the height of each end of the path, the length of the path, the nature of the terrain between the two points, the weather conditions along the path, the particular frequency used, etc.

Because of this variation, the propagation characteristics of the path between the transmitting and receiving antennas of a point-to-point communications system, or the paths between the antenna and possible targets in a radar installation, are of the greatest importance.

It is necessary to be able to predict the propagation characteristics of any given path, so that suitable paths can be chosen and equipment can be installed at sites that will permit the most reliable operation possible.

PATHS OF RADIO WAVES

Four different paths may be involved in the transmission of energy at radio frequencies from one point to another. These are the *direct, reflected, surface* and *sky waves.* In any given instance, one, two, three, or all four types may be present, all issuing from the same radiating source. This concept is suggested in Fig. 2-1 where only the direct and reflected waves are represented.

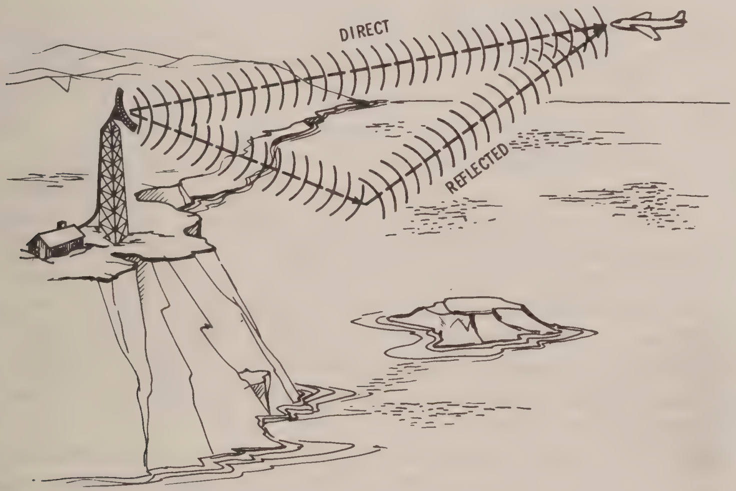


Fig. 2-1. Direct and reflected signals.

Most microwave installations make use of the direct wave only, although, in some cases, the reflected wave is also important. This characteristic of microwave systems may be contrasted with the behavior of similar installations operating in the VHF range, where the direct and reflected waves are usually (but not always) of equal importance.

The other two waves (surface and sky) are mentioned only because of their importance at low and medium frequencies. The surface wave is seldom important in either microwave or VHF operations, and the sky wave is absent at frequencies higher than 60 mc.

Direct Wave

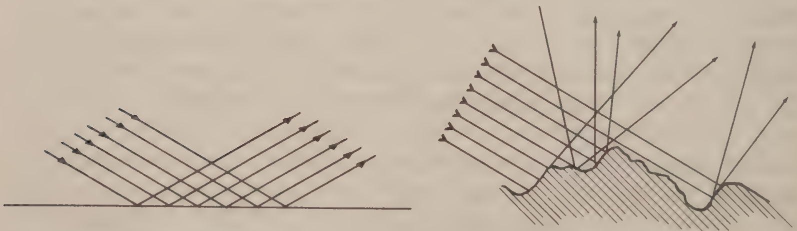
The direct wave from one point to another takes its name from the fact that it travels directly—that is, with no reflections—between two points. Fig. 2-1 shows the paths followed by the direct and the reflected waves traveling from the antenna of an early-warning radar to the aircraft that is under observation. As already noted, the success or failure of a microwave system usually depends on the presence or absence of the direct wave, since the other signals ordinarily are absent or unimportant. The first step in predicting the propagation characteristics of a given path is, therefore, consideration of the direct wave.

Reflected Wave

The term *reflected wave* is applied to a wave that has been reflected from the surface of land or water between the transmitting and receiving points. Fig. 2-1 shows the path of such a wave. There usually is a path of this sort. However, there can be *no* reflected signal unless the transmitting antenna actually radiates energy along a path which causes reflection. In point-to-point ground communications, the beamwidth of the antenna is such that ground reflections are always present.

A somewhat similar limitation holds for propagation in the reverse direction—from the aircraft to the antenna. Since the aircraft radiates an echo in *all* directions, a reflected signal always will be delivered to the antenna. This signal, however, has no effect on the performance of the system unless the antenna is able to pick up energy from the direction in which the wave arrives.

Another factor that affects the reflected wave is the nature of the reflecting surface. If the surface is smooth, any energy arriving along the reflected-wave path will be reflected in an orderly way; and at least a large part of it then will travel to


(A) Smooth surface.

(B) Rough surface.

Fig. 2-2. Radio wave reflection from different surfaces.

the target (Fig. 2-2A). If, however, the surface is rough, the energy that reaches it will be reflected in a multitude of different directions, and only a small portion of the energy will continue along the reflected-wave path to the target, as shown in Fig. 2-2B.

If the reflecting surface is rough, the reflected signal at the target will be so much weaker than the direct signal that it may be neglected. Generally speaking, at microwave frequencies water acts like a smooth surface, and land acts like a rough surface.

Although this discussion has been in terms of a target, (the aircraft shown in Fig. 2-1), the same considerations hold when the path ends at a receiving antenna.

The importance of the reflected wave lies in the fact that the effective signal at the target—or at the receiving antenna—is the sum of the direct and reflected waves. Like any other alternating waves, AC or RF, these signals must be added with due regard to the phase difference (if any) between them. A phase difference between the two waves often is introduced by the difference in the lengths of their paths. This difference may be anything from a small fraction of a wavelength to many wavelengths.

When the path difference is an odd number of half wavelengths, the two signals arrive at the target or receiving antenna in phase; this is true when the wave is reflected at small angles of incidence, causing a 180° phase reversal.

For horizontal polarization, the phase reversal is nearly 180°, irrespective of the frequency and magnitude of the grazing angle. This is true also of vertical polarization for almost all point-to-point ground communications. With vertical polarization at grazing angles in excess of, say 0.5° the phase reversal is less than 180°, depending on the frequency of the signal and the grazing angle.

Sky Wave

A sky wave is one that has been reflected from the ionosphere—a region that extends from an altitude of 40 or 50 miles to the upper limit of the atmosphere. The mechanism of ionispheric reflection of radio signals differs somewhat from that of ordinary reflection by a surface, described previously. Also, ionispheric reflection diminishes as the signal frequency increases, and therefore, it has no bearing on microwave propagation.

Surface Wave

The surface wave travels along the surface of the earth or water. It is not present unless the transmitting antenna is within a few wavelengths of the ground and, since it remains close to the surface, it cannot be picked up unless the receiving antenna is also low. The surface-wave signal is important at low frequencies (most daytime broadcast reception is by surface wave). However, it is relatively unimportant at VHF, and may be neglected completely at microwave frequencies.

CURVATURE OF MICROWAVE PATHS

The paths followed by microwave and VHF signals, whether direct or reflected, usually are slightly curved. The signals travel through the atmosphere at speeds that depend on temperature, atmospheric pressure, and the amount of water vapor present. Generally, the higher the temperature, the faster the signal; the lower the atmospheric pressure, the faster the signal; and the less amount of water vapor present, the faster the signal. The net result of these influences is that the signal speed changes with altitude, and, under normal atmospheric conditions, the variation is a small and uniform increase in the speed of the signal with increasing altitude. This causes all paths to curve slightly downward, as shown in Fig. 2-3.

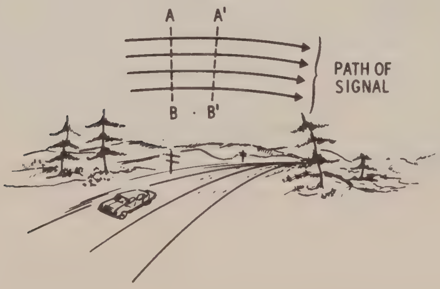


Fig. 2-3. Curvature of signal

To understand the reason for this curving, imagine a surface called a *wave front* such as the one represented by line AB where *A* is higher in altitude than *B*. A wave front is a surface, not necessarily flat, chosen in such a way that the phase of the signal is the same at all points on the surface. The signal always travels in a direction perpendicular to the wave front. Now, as wave front AB moves with the signal, it reaches position *A'B'* after a short time interval. The speed at *A* and *A'* is faster than the speed at *B* and *B'* since *A* and *A'* are at a greater altitude. Therefore, in a given time, the upper part of the wave front moves farther than the lower part, and the wave front tips slightly forward as it moves along. Since the path followed by the signal is always perpendicular to the wave front, it curves slightly downward as the wave front tips.

ATMOSPHERIC CONDITIONS

Reference has been made previously to normal atmospheric conditions. These are described here, and the more common abnormal conditions are described next.

Index of Refraction

Both light and radio waves travel through a vacuum with a speed nearly equal to 330 yards per microsecond. In more common units, they travel nearly 186,000 miles per second (these are statute miles of 5,280 feet). In almost any other media the velocity of the waves differs from the velocity in a vacuum, and signals of different frequencies travel with different velocities. Visible light, for example, travels through water about three-fourths as fast as through a vacuum, and radio waves travel through water only one-ninth as fast as through a vacuum. (Radio waves are, of course, greatly attenuated, or diminished, by water and, for practical purposes, it may be considered that they do not travel through water at all.)

Instead of saying that visible light travels through water with a velocity of about 247.5 yards per microsecond, it is more convenient to say that water has an index of refraction of about four-thirds for visible light. The index of refraction of water is simply a number equal to the velocity of light in a vacuum divided by its velocity in water. Similarly, water may be said to have an index of refraction of four-thirds for visible light, and nine for radio waves. A vacuum has an index of one for both types of signals. The velocity of a light or radio wave through any media is then nearly $330/n$ yards per microsecond, where n is the medium index of refraction for signals of the particular wavelength under consideration.

For microwave and VHF signals, the index of refraction of air becomes smaller with decreasing pressure, smaller with decreasing moisture content, and larger with decreasing temperature. All three of these variables—pressure, moisture content, and temperature—decrease with increasing altitude in a normal atmosphere. The changes in pressure and moisture content make the index smaller at greater heights. The change in temperature tends to make the index larger, but it is not sufficiently great to offset the decrease caused by reduction in pressure and moisture content. The net result, under average conditions, is that the index of refraction for microwave signals decreases uniformly with height up to altitudes of approximately 25,000 feet. Because the velocity of microwave signals increases with height, the signals follow paths that are curved slightly downward. Although the index of refraction for light waves also varies with height, it changes much less than the index for microwaves, and the paths followed by light waves are nearly straight lines.

Under normal atmospheric conditions the curvature in the paths of microwave signals is such that the paths appear as straight lines when they are plotted on a diagram in which the curvature of the earth's surface is distorted by using a fictitious value (equal to four-thirds of the actual value) for the earth's radius. Such atmospheric conditions are called *standard conditions,* and propagation (the process by which the wave travels, or the way in which it travels) under such conditions is called *standard propagation.*

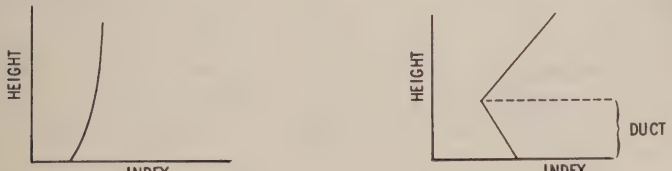
Anomalous Propagation

Temperature and moisture content are interrelated in such a manner that normal changes in these quantities do not appreciably alter the standard variation of the index with height. Because of this, standard conditions generally prevail. However, it is possible to have conditions under which the variation of the index with height is not standard. Such conditions are called *nonstandard conditions,* and propagation under such conditions is called *anomalous propagation.*

Light waves are refracted, or bent, when they pass from one medium to another having a different density. This physical fact accounts for the phenomenon known by travelers on the desert as a mirage. Thus, an oasis that is actually many miles away appears to be clearly above the horizon. What has happened is that light rays from the distant oasis have been bent around the earth's curvature. This effect may result when the atmosphere is not uniform and is subject to varying composition and density, causing the light rays to travel in a curved path instead of in a straight line. (Radio waves can behave in a similar fashion.) Both the optical and radio-wave mirage are caused by waves bending away from a medium of low density into a medium of higher density. The degree of bending is expressed by the changes in refractive index. The atmosphere may provide a curved path by which radio waves may be bent around the earth's curvature; such a curved path for radio waves is called a *duct.*

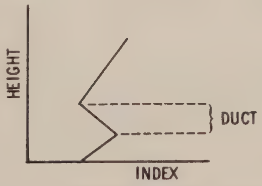
Atmospheric conditions of abnormal variation affect the operation of radio equipment on frequencies greater than 30 mc. In general, the higher the frequency, the greater will be the possibilities of a duct causing erratic radio-wave propagation effects. The radar operator using frequencies in the microwave region must be on guard against false readings attributable to atmospheric peculiarities. For this reason, an echo box is best for equipment performance testing instead of a target such as a cliff or a distant mountain. Such targets may not be dependable reference points, since abnormal propagation effects because of trapping can give false readings in signal strength as great as 50 db.

The question arises—what are the weather conditions that cause the creation of a duct and the resulting propagation peculiarities? It is well known that the earth's atmosphere is subject to many variations. In general, both temperature and moisture content decrease gradually and practically uniformly with height above the surface of the earth. This is the standard condition, and when this condition occurs, no ducts are formed, and radio transmission is normal. However, normal atmospheric conditions may not be present at all times. In some cases, the temperature first may increase with height and then begin to decrease. This is referred to as *temperature inversion.* In addition, the moisture content of the atmosphere



(A) Standard variation.

(B) Surface index.



(C) Sky duct.

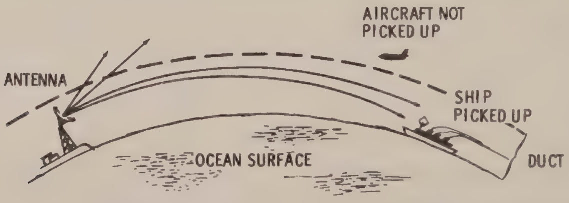
Fig. 2-4. Modified index of refraction curves.

may decrease very sharply close to the earth's surface. When these conditions exist—namely, temperature inversion accompanied by a region of great moisture variation—a duct, which can cause downward bending of radio waves, is likely to be formed.

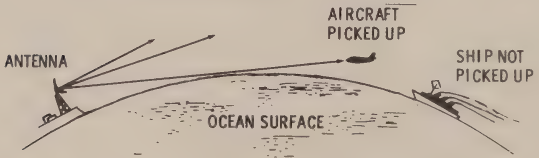
A modified index of refraction curve, usually called an M-curve (Fig. 2-4), is used to predict the probable degree of bending. Changes in M indicate the bending of the radio wave in the atmosphere. Fig. 2-4A shows a straight line sloping up gradually to the right without any break. A curve of this nature predicts standard propagation. Conversely, a curve that slopes up and to the left indicates that duct-forming conditions prevail. Fig. 2-4B shows an M-curve that rises in a direction toward the left, which signifies the possibility of a surface duct. In Fig. 2-4C the curve rises to the right and then breaks to the left. This discontinuity discloses the existence of atmospheric conditions that are likely to cause an elevated duct.

One of the weather conditions that causes anomalous effects results from warm moist air flowing over cold water; the air layers closest to the sea are cooled more than those of higher levels, and temperature inversion results. Since the upper levels are warmer, they can hold more moisture than the lower, colder levels, and therefore the vapor content increases with height. This weather situation may cause a duct to be formed and lead to nonstandard propagation.

Another weather situation that may cause the formation of a duct results from surface (radiation) inversions. This condition occurs when heat is being radiated from the earth's surface faster than it can be dispersed by the lowest level of



(A) Extended range by use of a duct.



(B) Normal range without use of a duct.

Fig. 2-5. Transmission with and without a duct.

the atmosphere. The result is that the temperature of the earth's surface is much lower than that of the air; and the temperature inversion which occurs may form a low-level duct. Radiation ducts may be expected over large, flat land areas in the middle and high latitudes. Desert country is such a region where this type of radiation duct may be encountered. Ducts of this character generally are found at night or in the early morning. In addition to the examples just cited, other weather situations can cause unusual changes in temperature and moisture distribution in the atmosphere. Two of these are the flow of cool air over a warmer sea and a descending or subsiding body of air which acts to bring about a warmer, dryer level.

As already mentioned, the paths of microwaves in a duct curve downward more sharply than usual. The presence of a duct always increases the operating range of a station (assuming that the range is limited by the curvature of the earth rather than by available power) and, when the curvature of a horizontal microwave path is the same as the curvature of the earth, the range may be extended by as much as several hundred miles, as shown in Fig. 2-5A. Fig. 2-5B shows the normal range of the same station. Both drawings are plotted on a fictitious earth curvature so that the signal paths in standard atmosphere, or above the duct, appear as straight lines. Although ducting increases the operating range of radar, it has a bad effect in that it may decrease the coverage. Fig. 2-5A shows that the high-altitude coverage of radar at long range has been reduced. In general, increased range is obtained when both target and radar lie in the duct, whereas the range is reduced when one is in the duct and the other is not. Because of this, an aircraft making an antishipping search by radar should always fly in a duct (when one is present) rather than above it. Since some energy is lost by the signal as it travels over the water, and some energy leaks out the top of the duct, extreme increases in range seldom are encountered except with ducts that are several-hundred wavelengths high.

Absorption and Scattering

The propagation characteristics of microwaves already discussed—curved paths and anomalous propagation—also hold for the VHF range of frequencies. Two additional atmospheric effects peculiar to microwaves are absorption and scattering.

Absorption—At certain frequencies the atmosphere will not transmit radio waves but will instead absorb most of the energy. The effect has nothing to do with reflection. Instead, the energy is absorbed in precisely the same way that red light is absorbed by a pair of green sun glasses (which passes only the green light). Certain of the gas molecules, particularly those of oxygen and water vapor, are able to oscillate at these frequencies and they behave exactly like resonant circuits that absorb energy only at the frequencies to which they are tuned. The effect is not present at frequencies lower than the X band and does not become serious until the K band is reached.

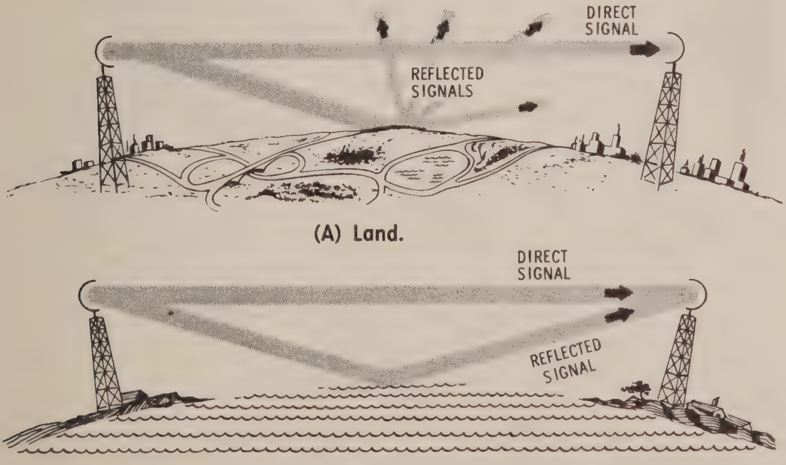
Scattering—The effectiveness of a microwave system may be reduced sharply by scattering of the signal. This phenomenon depends on the presence of small drops of water or fine particles of dust, snow, or ice in the air. These act as reflectors and divert some of the signal from its original path. The direction in which the signal is reflected has little relation to the original direction—except that it is not the same. The greater part of the energy is radiated in directions at right angles to the original path, and some is even reflected back in the opposite direction. As might be expected, the effect is worse at higher frequencies because of the greater size of the particles relative to the wavelength. Scattering does not affect S-band systems. It has a somewhat noticeable effect on X-band systems, however, and it has considerable effect on K-band systems. At K-band frequencies, the scattering effect of rain and fog is so great that the equipment may be useless in bad weather. It is possible to see storms with X- and K-band radars because a sizeable signal is returned to the radar by the scattering particles. The choice of operating frequencies, discussed in Chapter 1, is based on a compromise between improvement in fair-weather performance and increased vulnerability to adverse weather conditions, both of which accompany any increase in frequency.

IMPORTANCE OF REFLECTED SIGNAL

The reflected signal is important only when it is picked up at the receiving point with a strength comparable to that of signal may be neglected either because it is absent or because the direct signal. In most microwave installations the reflected it is weak as compared to the direct signal. When the reflected signal is present and strong, its importance lies in the fact that *it may either reinforce the direct signal or cancel it almost completely*. Whether the two signals tend to reinforce or cancel each other depends on their phase difference. As noted earlier, this phase difference results from the different path lengths of the direct and reflected signals. When the two signals are in phase with each other, or nearly so, they may be added directly. If they are of nearly equal strength, the combined signal is twice as strong as either one alone, and it has a power level four times as great. But, when they are close to 180° out of phase with each other, *there will be partial cancelation, since the reflected signal (which is usually the weaker) subtracts from the strength of the direct signal; when the two signal levels are equal, or nearly so, cancelation may be substantially complete.* It is desirable, therefore, to consider a number of representative installations with a view to determining when the reflected signal must be considered and when it may be neglected. Among the determining considerations are—whether the transmission path is over land or water; whether the beam is highly compressed in height or width or both; whether the target (or receiving antenna) is moving or stationary; whether the signal is polarized horizontally or vertically; and the angle at which the signal strikes the reflecting surface. These considerations are discussed in the following paragraphs.

Overland and Overwater Operation

It has been pointed out that land is a rough surface and therefore a poor reflector at microwave frequencies; whereas water is a good reflector. Fig. 2-6A shows a point-to-point communications installation having an overland transmission



(A) Land.

(B) Water.

Fig. 2-6. Reflected signals from land and water.

path; and Fig. 2-6B shows a similar installation with an overwater transmission path. In Fig. 2-6A the reflected signal is radiated in all directions from the point of reflection. Only a small portion of the energy reaches the receiving point, and the reflected signal may be neglected. In Fig. 2-6B, however, the water surface acts as a mirror, and the reflected signal is of nearly the same strength as the direct signal.

In general the reflected signal is unimportant and consequently may be neglected when the transmission path lies over land. Although the reflected signal is obviously important in the case of the overwater point-to-point communications system, it need not be necessarily considered in *all* overwater microwave installations.

TRANSMITTING-ANTENNA DIRECTIVITY

Tracking Radars

A tracking radar transmits and receives with an antenna having a highly directive pattern. The beam is compressed in both height and width and is like a searchlight beam. Although it is always possible to draw the path of a reflected signal between the radar and the target, only a negligible amount of energy travels along that path.

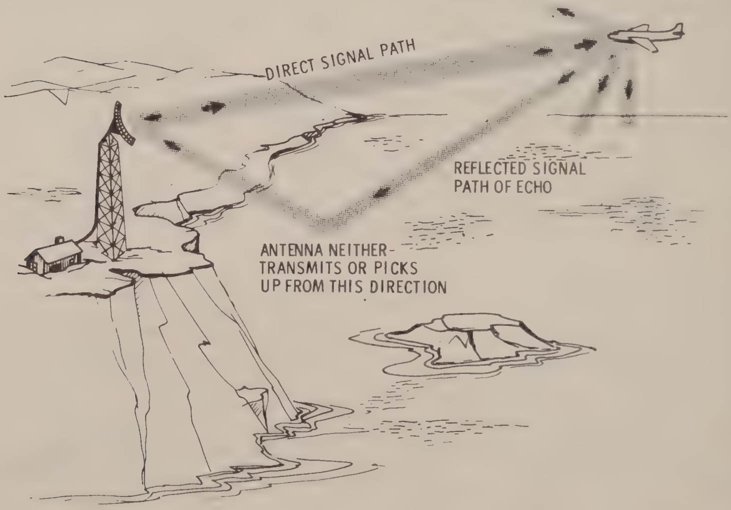


Fig. 2-7. Direct and reflected signal paths for tracking radar.

Fig. 2-7 shows a tracking radar pointing at an aircraft. Because of the difference in the directions of the direct and reflected paths, little or no reflected signal is transmitted to the target. The target, of course, returns echoes in all directions. In addition to the echo arriving along the direct path, an echo signal arrives at the radar by way of the reflected-signal path. This, however, is not picked up because the antenna is highly directional for both transmission and reception.

If both the tracking radar and the target are close to the water surface, it is possible to have a situation such as that shown in Fig. 2-6B, where the direct and reflected paths are in nearly the same direction. Here, the reflected signal is important in the case of a point-to-point communication system.

In the tracking radar, however, the target is moving rapidly and the direct and reflected signals cancel at one instant and add at another. The average effect is nearly the same as though the reflected signal were absent, and consequently, it may be neglected. It is important in the communications system only because both transmitter and receiver are stationary so that the effect, whether cancelation or addition, is constant.

Search Radars

The antenna of a search radar has a pattern that is very sharp in the horizontal direction and fairly broad in the verti-

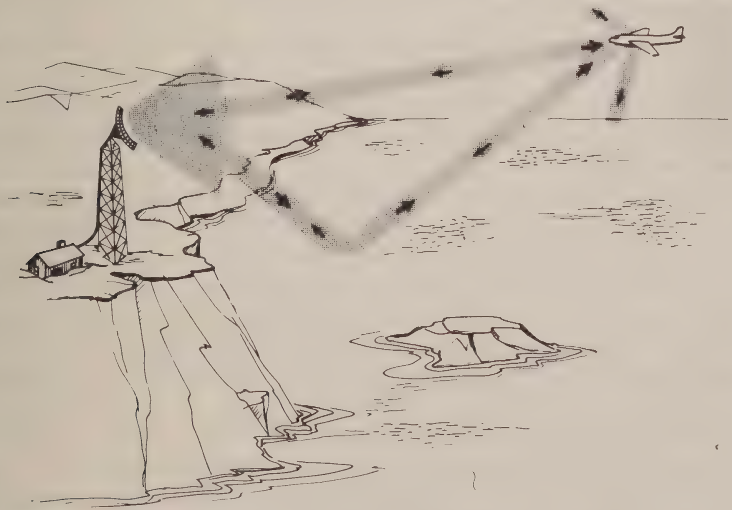


Fig. 2-8. Direct and reflected signal paths for search radar.

cal plane (Fig. 2-8). The result is that the reflected signal must always be considered in a shipborne search radar or a ground-based search radar that looks over water.

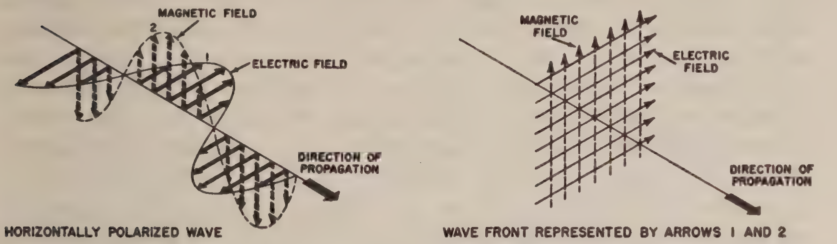
The reflected wave is present, as well, in an airborne search radar operated over water. The target, however, is relatively large—a ship or a coast line—so that cancelation of the direct and reflected signals at one point on the target is likely to be accompanied by addition of the signals at another point. The effect of the reflected signal, therefore, is usually neglected.

HORIZONTALLY AND VERTICALLY POLARIZED WAVES

Since land is rough at microwave frequencies, it reflects signals in all directions in much the same manner that a white painted wall reflects light; however, water is smooth, and it reflects signals in the same manner as a mirror reflects light. There is, of course, a great deal of reflection from a land surface—every airborne radar operator has seen ground returns, or echoes from the ground. However, in the case of transmission over land between two points, the reflected signal is so weak compared to the direct signal that its effects may be disregarded. At microwave frequencies it is necessary to consider the reflected signal only in an overwater point-to-point communications installation, a shipborne search-radar installation, and a ground-based search-radar installation which looks over water. The remainder of this discussion, therefore, is concerned only with the reflection of microwaves from a water surface.

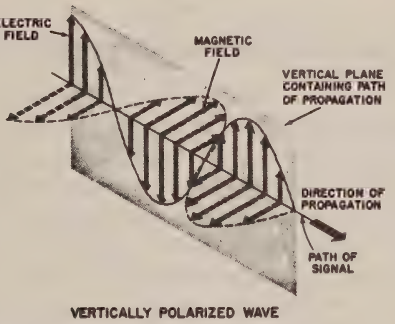
It is possible to distinguish between two types of waves, called horizontally and vertically polarized waves. The significance of this distinction to the present discussion lies in the different behavior of the two signals on reflection from a water surface.

A radio wave consists of both an electric and a magnetic field, each related to the other in a definite way. Both exist in the same region of space, and both move together at the same velocity in the direction of propagation. At any one point and any one instant each field has a definite direction and strength. Along the direction of propagation at any one instant each field varies in strength and alternates in direction. Fig. 2-9A represents this variation. Each solid arrow represents the direction and strength of the electric field throughout a surface called a wave front which is perpendicular to the direction of propagation. The strength of the field throughout the wave front is proportional to the length of the arrow; the direction of the field is the same as that of the arrow. Similarly, each broken arrow represents the direction and strength of the magnetic field through such a surface. Fig. 2-9B shows a small part of one wave front of the wave represented in Fig. 2-9A. The entire field contains an infinite number of these wave fronts, all moving in the same direction and at the same velocity.


MAGNETIC FIELD
2
1
ELECTRIC FIELD
DIRECTION OF PROPAGATION
HORIZONTALLY POLARIZED WAVE
MAGNETIC FIELD
ELECTRIC FIELD
DIRECTION OF PROPAGATION
WAVE FRONT REPRESENTED BY ARROWS 1 AND 2

(A) Horizontally polarized.

(B) Representation of wavefront.


ELECTRIC FIELD
MAGNETIC FIELD
VERTICAL PLANE CONTAINING PATH OF PROPAGATION
DIRECTION OF PROPAGATION
PATH OF SIGNAL
VERTICALLY POLARIZED WAVE

(C) Vertically polarized.

Fig. 2-9. Horizontally and vertically polarized waves.

As the wave fronts sweep past a fixed point in the path of the wave, both the electric and the magnetic fields at that point alternate at the frequency of the signal, just as the voltage and current in an AC circuit alternate at the generator frequency.

At distances greater than several wavelengths from the signal source the two fields are in phase with each other. That is, the electric and magnetic fields at a particular point reach maximum strength at the same instant. Also at a particular instant, they both have maximum strength at the same point, as in Fig. 2-9A.

At every point in the path of the wave—that is, at every point on every wave front—and at every instant, the electric and magnetic fields are perpendicular to each other, and both are perpendicular to the direction of propagation.

A horizontally polarized wave is one in which the direction of the electric field is horizontal, as in Fig. 2-9A. A vertically polarized wave is one in which the direction of the electric field lies in a vertical plane containing the path of the signal, as in Fig. 2-9C. It is not necessary that the electric field be vertical, but only that it lie in a vertical plane. Thus, if the signal path of the wave represented in Fig. 2-9C is inclined downward, the electric field is not vertical, but the wave nevertheless is polarized vertically. As a matter of interest, the plane containing the direction of the electric field also contains the radiating element. Thus, a horizontal antenna radiates horizontally polarized signals, and a vertical antenna radiates vertically polarized signals.

Reflection of Horizontally Polarized Signals

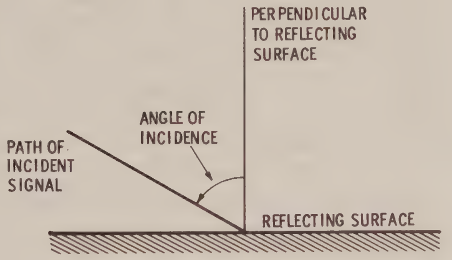
When a horizontally polarized wave strikes a horizontal surface, the electric field is parallel to the surface. This causes current to flow in the surface, and the current radiates the reflected signal. Each portion of the surface radiates reflected signals in all directions, but because of phase differences, the reflected signals from various portions of the reflecting surface cancel each other in all directions except the one taken by the reflected wave.

The incident wave (the wave falling on the surface) and the reflected wave differ in phase by 180° at the surface, so that one nearly cancels the other. The small electric field remaining after cancellation is responsible for the current flow in the surface. The greater the conductivity of the surface, the smaller is the electric field required to cause current flow, and the more nearly equal is the strength of the reflected wave to that of the incident wave. Sea water is a good conductor at the microwave frequencies, and it is usual to say that the reflected signal has the same strength as the incident signal. A horizontally polarized signal therefore is considered to be reflected from a water surface with a phase reversal (a phase shift of 180°) and with a negligible reduction in level.

Reflection of Vertically Polarized Signals

The reflection of vertically polarized signals from the surface of the sea is much more complicated than that of horizontally polarized signals. Both the phase and the level (strength) of the reflected signal change with the angle between the surface and the direction of the signal. In addition, the mechanism of reflection is more complicated than it is for horizontally polarized signals. The direction of the incident signal may be specified by an angle called the *angle of incidence*. This is the angle between the path of the signal and a line perpendicular to the reflecting surface (Fig. 2-10).

Fig. 2-10. Angle of incidence of a vertically polarized signal.



When a vertically polarized signal strikes the surface of the sea at an angle of incidence close to 90°, the reflected signal is almost as strong as the incident signal. As the angle of incidence decreases, the reflected signal becomes weaker until it is only about one-fifth as strong as the incident signal at an angle of 84°; with further decrease in the angle, however, the reflected signal becomes stronger again.

When the angle of incidence is nearly 90°, the vertically polarized signal also suffers a phase reversal on reflection, but the phase change decreases only 90° when the angle of incidence is 84° and continues to decrease as the angle of incidence is reduced further. These figures apply to the S-band, but the same effects are present at both X- and K-bands, although the figures are different for each band. The phase angle, or reflection of vertically polarized waves for a given angle of incidence, increases with an increase in frequency. Consequently, microwave installations usually operate with horizontally polarized signals when the reflected signal is of importance.

EFFECT OF WAVE PATHS ON COMMUNICATIONS INSTALLATIONS

It has already been pointed out that the reflected signal may be neglected in an overland communications installation. The effect in an overwater installation is shown in Fig. 2-6B. When the direct path nearly touches the surface of the sea, the direct and reflected paths have the same length. Since the reflected signal suffers a phase reversal on reflection, it will arrive at the receiver precisely 180° out of phase with the direct signal, and at nearly the same strength. The result will be almost complete cancellation, and reliable operation will be impossible.

If either or both antenna is made higher, the direct path clears the surface of the sea, and the length of the reflected path is increased. If the height increase is chosen correctly, the reflected-path length will be ½ wavelength longer than that of the direct path. This causes a phase difference of 180° between the direct and the reflected signals. The phase difference resulting from the difference in path lengths is added to the 180° phase difference which occurs because of the phase reversal on reflection, and the two signals will arrive in phase and add to each other. Additional increases in height cause further increases in the length of the reflected path, and this in turn results in successive additions and cancellations of the signals. The calculation of the best antenna heights for a given installation is discussed subsequently.

EFFECT OF WAVE PATHS ON SEARCH-RADAR INSTALLATIONS

The effect of the reflected signal on the operation of a search radar is straightforward and can be understood easily by referring to Fig. 2-11. The radar antenna is located at *A*, and the direct path extends toward the target in direction AB. The reflected signal reaches the water at *C* and then is reflected in direction CD. Since the target is far away; directions AB and CD are essentially the same. The effect is as though another radar were located at *E*, a point just as far below the reflecting surface as the actual radar at *A* is above the surface. Distance AE is then equal to 2H, where *H* is the height of the radar antenna above the reflecting surface. Triangle AEC is an isosceles triangle, in which distance *EC* is equal to distance AC. Therefore, the length of the reflected path is not altered if it is assumed that the reflected signal originates at the image radar located at *E*.

Because of the phase reversal on reflection the signals from the image radar must be precisely 180° out of phase with those from the real radar. The two signals at the aircraft are then the same as though they were received from the rear radar and its image. Next AF is drawn perpendicular to path *ED*. Since AB and FD are parallel, the distance to the aircraft is the same, whether measured from *A* or *F*. The geometrical difference in the two path lengths then is EF.

Because the two transmitters are 180° out of phase, the direct and reflected signals will add at the target when EF is an odd number of half wavelengths and will cancel when EF is an even number of half wavelengths. Therefore there are certain directions in which the radar cannot pick up a target, except at extremely short ranges.

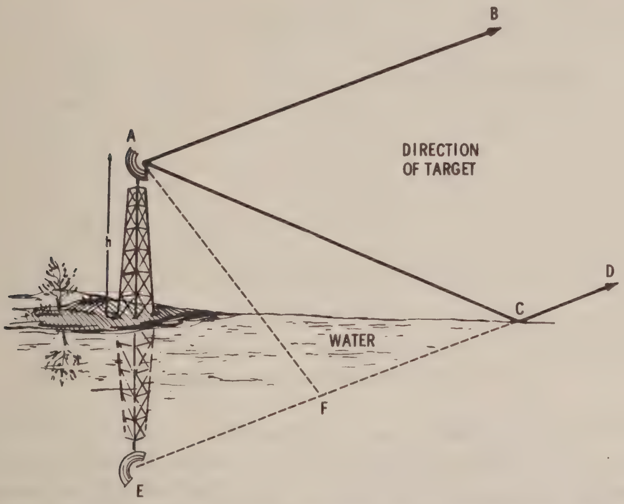


Fig. 2-11. Signal-path difference for search radar.

Vertical Polar Diagram

The combined radiation pattern of the actual radar and its image is the result of alternate addition and cancellation of the direct and reflected signals. It may be plotted on a height-range diagram to show the coverage of the radar. Such a plot, called a vertical polar diagram, is shown in Fig. 2-12. The parts of the pattern in which the radar has good performance (the directions in which the signals add) are called *lobes*, and the parts in which the radar cannot pick up targets are called *gaps*.

The VPD (vertical polar diagram) in Fig. 2-12 shows, for a radar located 1,000 feet above sea level, gaps extending from 70 to 90 miles at 15,000 feet, from 100 to 135 miles at 10,000 feet, etc. There is always a gap at the horizon, because both path lengths are the same in that direction. As the wavelength at which the system operates is decreased, the lobes and gaps become narrower and closer together. This

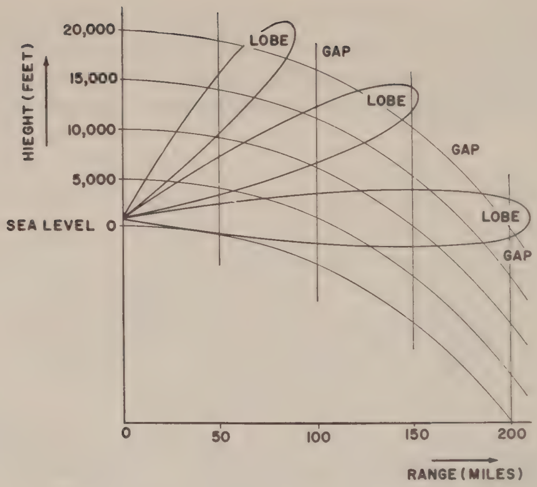


Fig. 2-12. Vertical polar diagram.

results from the fact that a smaller change in direction is required for a given change in the phase difference between the signals at the target (Fig. 2-11). A given change in direction makes a given change in distance EF. Such a change, however, may be only a fraction of a wavelength at very high frequencies or several wavelengths at microwave frequencies. In practice the separation of the radar and its image, as measured in wavelengths, is the determining factor. If distance AE is several hundred wavelengths, a small change in direction is still able to change *EF* by several wavelengths.

The VPD of a VHF radar shows relatively wide gaps, because the height of the radar may be as little as 8 or 10 wavelengths (the height is one-half of AE). At microwave frequencies, however, the height of the radar is at least 100 wavelengths, and may be several thousand. The result is that the gaps are so narrow and the lobes are so closely spaced on the VPD that the general effect is one of uniform vertical coverage. In most cases, therefore, the effect of the reflected signal is simply to increase the apparent power of the search radar. It must be emphasized, however, that this is not true of VHF radars.

LINE-OF-SIGHT LIMITATIONS

The path followed by a light wave in traveling from one point to another is called the optical line of sight between the two points. The path followed by a microwave or VHF signal under standard atmospheric conditions is called the radio line of sight or, less frequently, the microwave or VHF line of sight between the two points. For most purposes optical lines of sight may be considered as straight lines. Radio lines of sight, except when they are vertical, curve slightly downward.

It has been explained that the effects of the reflected signal can be neglected at microwave frequencies, except in an overwater point-to-point system. This is true even in the case of an overwater search radar, where the level of the reflected signal ordinarily is comparable to that of the direct signal. In an overwater point-to-point communications system successful operation requires the existence of a suitable path for the direct signal and suitable phase relations between the direct and reflected signals at the receiving point. In all other cases, successful operation requires nothing more than a suitable path for the direct signal. (It is assumed, of course, that suitable equipment with adequate power output for the installation is available.)

Under conditions of standard propagation the direct wave follows a radio line of sight. Usually, therefore, it is necessary only to determine whether an unobstructed radio line of sight connects the two points between which microwave operation is desired.

Overland Point-to-Point Communications System

As stated in the preceding paragraph, successful operation of a microwave overland point-to-point communications system (assuming adequate equipment) under conditions of standard propagation requires nothing more than the existence of an unobstructed radio line of sight connecting the two points between which operation is desired. Two ways of determining whether such a line of sight exists are by visual inspection of the path, or by construction of a profile diagram.

Inspection of Path—In many cases, particularly where the two points are not too far apart, it is possible to make a direct

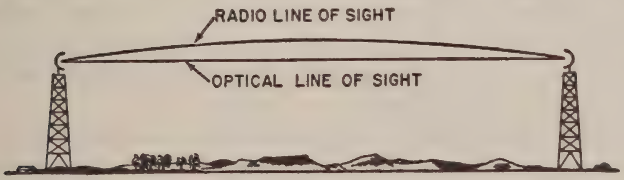


Fig. 2-13. Optical and radio lines of sight.

inspection of the path. If the proposed position of one antenna can be seen (optically) from the proposed position of the other antenna, satisfactory operation will usually be possible. The optical line of sight joining the two antenna positions is a straight line, whereas radio line of sight curves downward. Consequently, the radio line of sight is above the optical line of sight at all points, except the antenna positions, where the two lines of sight coincide (Fig. 2-13). If, therefore, there are no obstructions along the optical line of sight and no obstructions above it, the radio line of sight will be clear. It is usual to insist that there be no obstructions within 50 feet or so of the optical line of sight.

The statement that the radio line of sight curves downward, but is above the optical line of sight may be misleading to some. When transmitted, radio waves tend to curve toward the earth's surface as they travel along. Thus, they curve down; however, when aiming two point-to-point antennas toward each other, they are each aimed high. That is, if one were to follow the optical line-of-sight from one antenna to the other, he would actually sight over the one he was looking to from the one he was looking from. This aiming high allows reception at further distances.

This can be compared to a ball player throwing a pass. If an outfielder were to throw to the infield, and throw the ball *exactly parallel* with the ground, the ball would not get far; it would hit the ground. Instead, he aims high. The ball then curves downward as it is traveling upward. In this manner, the ball travels further—in an arc from the thrower to the receiver. While the ball is said to have been thrown upward; at the instant it leaves the hand it starts curving down due to the pull of the earth.

In the same manner, the radio wave is pulled toward the earth, due to the earth's influence (magnetic field). Therefore it is said that while the wave curves downward, it is actually above the line of sight.

Profile Diagram—When it is impossible to inspect the path between two points, or when the results of an inspection are not clear, it is necessary to make a *profile diagram* of the path in order to decide whether there is an unobstructed radio line of sight for the direct signal. A profile diagram is nothing more than a picture that shows the height of the land along a path between two points.

Consider, for example, two 100-foot antenna towers that are separated by 20 miles of smooth, flat country. Fig. 2-14A shows a profile diagram based on a flat earth; that is, the earth is assumed to be flat. The size of the diagram, of course, has been reduced; originally, it was drawn with a horizontal scale of 2 miles per inch and a vertical scale of 100 feet per inch. The reason for the different horizontal and vertical scales will be discussed later. Since the diagram is based on a flat earth, the surface of the earth is indicated by a straight line.

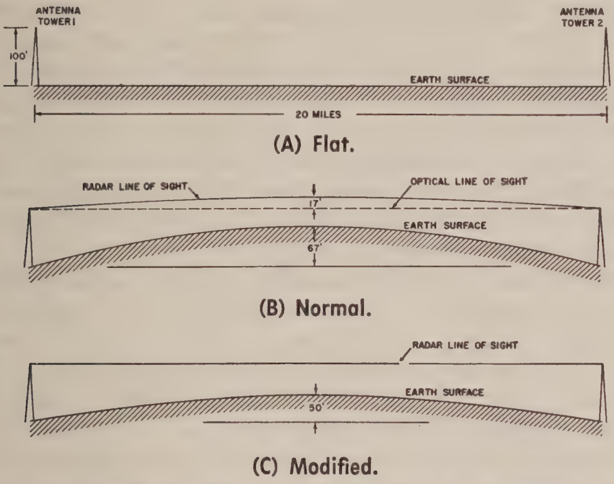

ANTENNA TOWER 1
ANTENNA TOWER 2
100'
EARTH SURFACE
20 MILES
(A) Flat.
RADAR LINE OF SIGHT
OPTICAL LINE OF SIGHT
17'
EARTH SURFACE
67'
(B) Normal.
RADAR LINE OF SIGHT
EARTH SURFACE
50'
(C) Modified.

Fig. 2-14. Radio line of sight over flat, normal, and modified earth.

Fig. 2-14B shows the same antennas but takes account of the curvature of the earth. Referring to Appendix I, it is seen that the surface of the earth between two points rises a distance that is equal to $S^2/6$ feet above a straight line (through the earth) between the points, where S is the distance in statute miles between the points. For the present instance, S is 20 miles, and $S^2/6$ is 67 feet. The surface of the earth is shown by an arc of a circle, and the arc is chosen so that its midpoint is 67 feet above the straight line between the tower bases. (The distance of 67 feet, like the heights of the antenna towers, is measured on the scale of 200 feet per inch.)

Since this profile diagram takes into account the curvature of the earth, any optical line of sight is a straight one. The radio line of sight between the antennas curves downward, and therefore lies above the optical line of sight. The maximum distance between the two lines of sight is shown in the diagram as 17 feet, and may be found by the following method.

For working with microwaves it is desirable to use a profile diagram based on a modified earth. On such a diagram, the radio line of sight, not the optical line of sight, appears as a straight line. The modified earth has a radius that is four-thirds of the radius of the real earth (4,000 miles) and its surface therefore is curved somewhat less. It is shown in Appendix I that the surface of the modified earth rises only $S^2/8$ feet above a straight line between two points separated by distance *S*. Fig. 2-14C shows a profile diagram based on the modified earth. As before, the surface of the earth is shown by an arc of a circle, but the midpoint of the arc is now only $S^2/8$ feet (50 feet) above the line joining the bases of the antenna towers. This distance differs from the corresponding distance of 67 feet on the normal earth diagram shown in Fig. 2-14B by 17 feet—the true maximum distance between the radio and the optical lines of sight.

The construction of the modified earth profile diagram in Fig. 2-14C is equivalent to a distortion of the diagram in Fig. 2-14B, which makes the downward-curving earth surface a little less curved—by precisely the right amount to distort the curved radio line of sight into a straight line. Therefore, as mentioned previously, all radio lines of sight drawn on a modified earth profile diagram are drawn as straight lines.

As an example of the use of a profile diagram, consider the following problem: Two relay units are to be installed as part of a microwave point-to-point communications system. Each relay unit is designed to receive signals and pass them on. Therefore, it is apparent that reliable operation between

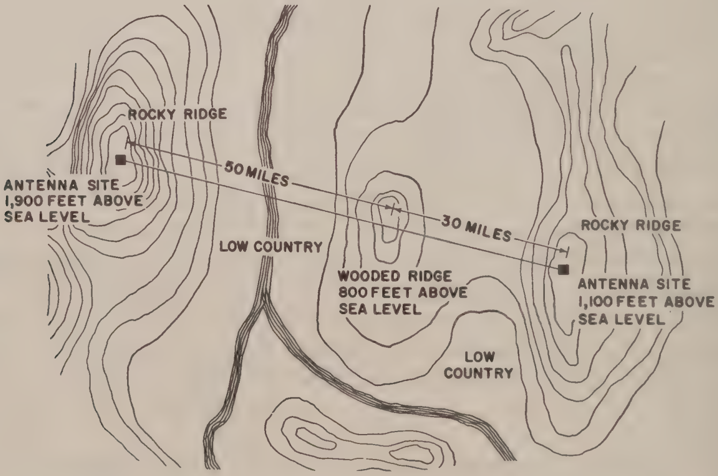

ROCKY RIDGE
50 MILES
ANTENNA SITE
1,900 FEET ABOVE
SEA LEVEL
LOW COUNTRY
30 MILES
ROCKY RIDGE
WOODED RIDGE
800 FEET ABOVE
SEA LEVEL
ANTENNA SITE
1,100 FEET ABOVE
SEA LEVEL
LOW
COUNTRY

Fig. 2-15. Map of profile-diagram problem.

the two units will be necessary if the communications system is to serve its purpose. Proposed sites have been chosen for the units, as shown on the map in Fig. 2-15. One relay unit is to be located on top of a rocky ridge 1,900 feet above sea level, and the other is to be located on a similar ridge 1,100 feet above sea level. The country between the two locations is low except for a wooded ridge 800 feet above sea level, 50 miles from the higher location. Assuming that the antenna towers are to be 100 feet high, the antennas will be 2,000 feet and 1,200 feet above sea level. The problem is this—can the two relay units operate successfully across the wooded ridge between them, or will a third relay unit be required?

The problem is easily solved by referring to the profile diagram in Fig. 2-16, which is constructed as follows: Lay out horizontal line AB and vertical line CD, making sure that they are perpendicular to each other. Using a suitable horizontal scale (40 miles to the inch in this diagram), lay out the positions of the antenna locations and the wooded ridge which is under consideration as a possible obstruction. All horizontal distances are measured from the center of the path. The site of the higher antenna is therefore 1 inch to the left of *E* (the intersection of the horizontal and the vertical lines) or 40 miles from the center of the path, at *F*. The lower antenna is at *G*, an equal distance on the other side of *E*. The wooded ridge is at *H*, ¼ inch, or ten miles, from *E*.

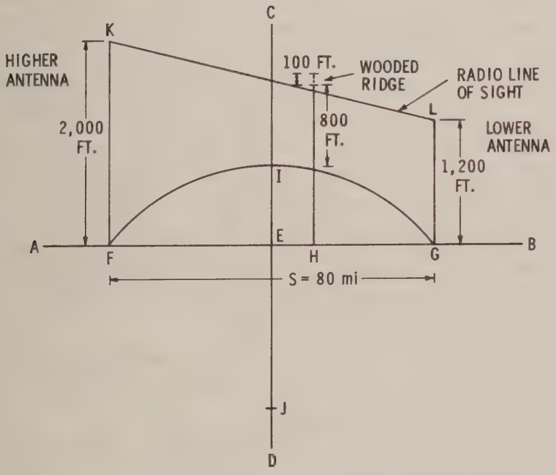


Fig. 2-16. Profile diagram of a height-range problem.

Since *S*, the distance between the two antennas is 80 miles, the height of the midpoint of the arc is 800 feet (using $S^2/8$, which applies to the modified earth). Choosing a convenient vertical scale, say 1,600 feet to the inch, locate *I*, which is ½ inch, or 800 feet, above *E*. Next locate *J*, which is the center of a circle passing through *F*, *G* and *I*, and draw the arc through these points. *J* is located by trial and error. Place the point of a compass somewhere on the vertical line below *E*, set the radius so that the circle will pass through *I*, and see whether the same circle passes through *F* and *G*. If it falls outside these points, select a new center higher up, set a new and shorter radius, and try again. If a circle is finally obtained which passes through *F*, and *I* but not through *G*, or through *G* and *I* but not through *F*, the original lines AB and CD, are not perpendicular and it will be necessary to start again.

Having drawn the arc, locate the higher antenna at *K* by measuring upward 1.25 inches, or 2,000 feet, from *F*. Note that this measurement is made in a direction perpendicular to AB, *not* in a direction perpendicular to the arc. Locate the lower antenna at *L* by measuring upward 0.75 inches, or 1,200 feet, from *G*. The top of the wooded ridge, *N*, now is drawn in. It is located directly above *H*, and its height is ½ inch, or 800 feet, above *M* on the arc that represents the surface of the earth. The line upward from *H* may be extended an additional 100 feet to allow for the height of the trees. This has been done on the drawing and the extension is shown as a dotted line. Finally, draw the radio line of sight as a straight line between the two antennas, *K* and *L*.

It is clear from this profile diagram that the two relay stations will not operate properly unless a third unit is located on the top of the intervening ridge.

Profile-Diagram Error—The profile diagram shown in Fig. 2-14C is based on a horizontal scale of 40 miles to the inch, or 211,200 feet to the inch. It is not practical to use the same scale for vertical distances since the midpoint of the arc is only 800 feet above the base line, and this would be immeasurable on the diagram. When the vertical scale is expanded—in this diagram, 1,600 feet per inch—some error is introduced. The error is a height error; that is, the arc that represents the surface of the earth is not exactly where it should be. The error is zero at the center of the arc and at each end, and reaches its greatest value at two points 0.15(S) from each end. In the problem example given, *S* is 80 miles, and the maximum error occurs at points 12 miles from each antenna.

The maximum error in feet is $(S^2/8)\ (H/W)^2$, where S is the distance in miles between the points on which the profile diagram is based, H is the midpoint height of the arc in inches on the diagram, and W is the width of the arc in inches on the diagram. In the example above, S is 80 miles, so that $S^2/8$ is 800 feet. H is ½ inch (800 feet on a scale of 1600 feet per inch), and W is 2 inches (80 miles on a scale of 40 miles per inch). Quantity (H/W) then is equal to ¼ and $(H/W)^2$ is 1/16. The maximum error, therefore, is 800/16, or 50 feet. The error at the point of interest is, of course smaller, since the maximum error occurs at points only 12 miles from the antennas. If it is necessary to reduce the error, a different vertical or horizontal scale can be used so that H/W becomes smaller.

Radar Coverage

In considering a proposed point-to-point communications installation it is necessary only to predict whether operation will be possible between the proposed sites. In a radar installation, however, it is necessary to predict the coverage available from a proposed site—that is, the ranges at which various targets may be detected. In the VHF range of frequencies, coverage may be affected by gaps, although there are a number of methods of filling the gaps. At microwave frequencies, the vertical polar diagram is effectively free from gaps, and with adequate equipment the coverage is limited only by obstructions which interfere with the radio line of sight to the target.

Height-Range Diagrams—Radar coverage usually is shown on a height-range diagram such as that in Fig. 2-17. Height-range diagrams are similar to profile diagrams; and they are based on the same mathematics. The only important difference is that the equipment site is shown at the midpoint of the arc on a height-range diagram, because the maximum error in the diagram occurs near the longest ranges instead of near the shortest ones. Since coverage is often different in different directions, a separate height-range diagram must be prepared for each direction that is of particular interest. A height-range diagram usually shows coverage in a single direction and is similar to half of a profile diagram.

Referring to Fig. 2-17, a height-range diagram may be constructed in the following steps:

1. Draw horizontal line AB and vertical line CD, making sure that they are perpendicular to each other. Although

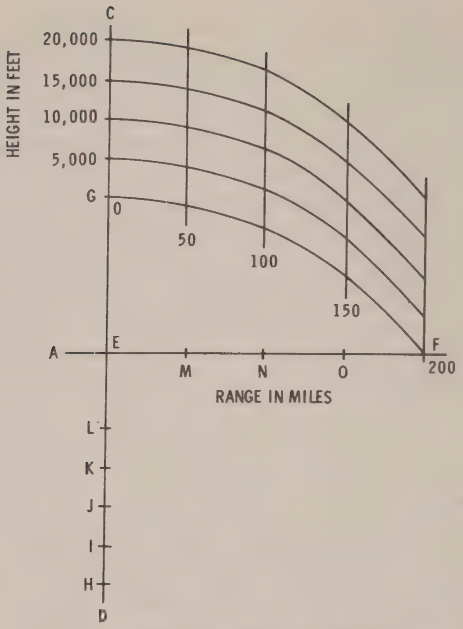


Fig. 2-17. Radar height-range diagram.

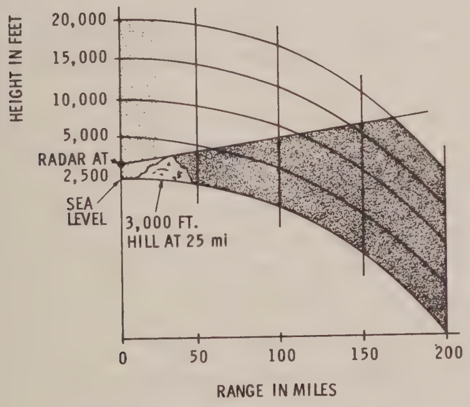
any convenient distance may be taken as the basis for a height-range diagram, 200 miles is probably a usual choice. In Fig. 2-17, a horizontal scale of 100 miles to the inch was used, and *F* is located 2 inches (200 miles) to the right of *E*, the intersection of lines AB and CD.

2. Since EF, which is 200 miles, is only half of the base line of the equivalent profile diagram, the height of the arc is calculated by using 400 miles for *S* so that $S^2/8$ is 20,000 feet. A vertical scale of 20,000 feet to the inch has been used in Fig. 2-17, and the next step is the location of *G*, which is 1 inch, or 20,000 feet, above *E*.
3. Next find *H* (on *CD*) which is the center of a circle passing through *F* and *G*, and draw in arc GF, which represents the surface of the earth. *I, J, K,* and *L* are located at ¼-inch (5,000-foot) intervals above *H*, and these intervals are used as centers for the 5,000-, 10,000-, 15,000-, and 20,000-foot arcs that are all drawn with the same radius used for arc GF.
4. Finally, *M, N,* and *O* are located at intervals of ½-inch, or 50 miles, and the vertical range lines are drawn in parallel to CD.

The maximum height error occurs at a range of 0.15 (S) to the left of point *F*. Since *S* is 400 miles, 0.15 (S) is 60 miles, and the greatest error is at a range of 140 miles. The size of the maximum error is $(S^2/8)$ $(H/W)^2$, where S, H, and W have the values associated with the equivalent profile diagram. Thus, S is 400, and $(S^2/8)$ is 20,000 feet. H is 4 inches, W is 16 inches (twice the width of the height-range diagram,) and therefore, $(H/W)^2$ is 1/16. Consequently, the maximum height error, which occurs at a range of 140 miles, is 1,250 feet. The error drops to zero at ranges of zero and 200 miles (the midpoint and end of the arc.) If a vertical scale of 40,000 feet to the inch is used (or if the horizontal scale is changed to 50 miles to the inch,) the maximum error will drop to 312.5 feet.

Radar Coverage Over Land

Fig. 2-18 illustrates the method of showing radar coverage on a height-range diagram. This diagram is made for a radar sited 2,500 feet above sea level. The only significant obstruction in the direction for which the diagram is made is a hill 3,000 feet high at a range of 25 miles. The profile of the hill is plotted in the manner described in the discussion covering Fig. 2-16, the height being measured vertically upward from the arc representing the surface of the earth.



MINIMUM RANGE 20 MILES
MAXIMUM RANGE 57 MILES ON TARGET AT 5,000 FT
110 MILES ON TARGET AT 10,000 FT.
143 MILES ON TARGET AT 15,000 FT.
170 MILES ON TARGET AT 20,000 FT.

Fig. 2-18. Radar-coverage graph.

A radio line of sight is drawn as a straight line on the diagram, from the transmitter through the crest of the hill and beyond. Below this line, in the shaded region, there is no coverage under standard conditions because of the obstruction posed by the hill. The shaded region at the left, at ranges shorter than 20 miles, is a region of poor coverage. Echos from targets in this region may be obscured by echoes from the ground. The clear region, above the limiting radio line of sight and at ranges greater than 20 miles, is the region of good coverage.

Radar Coverage Over Water

A much simpler problem is posed when most of the line of sight lies over water. A height-range diagram seldom is used, because the maximum range on a target at a given altitude can be calculated easily.

The geometry of the problem is shown in Fig. 2-19A. The maximum range, *S*, is the sum of S_1 and S_2, where S_1 is the distance from the radar to the horizon and S_2 is the distance from the horizon to the target. In order to find *S*, then, it is necessary only to know the law that determines the distance to the horizon when the height of a point is known. The problem is reduced to the solution of the right triangle, shown in Fig. 2-19B. By Pythagoras' theorem,

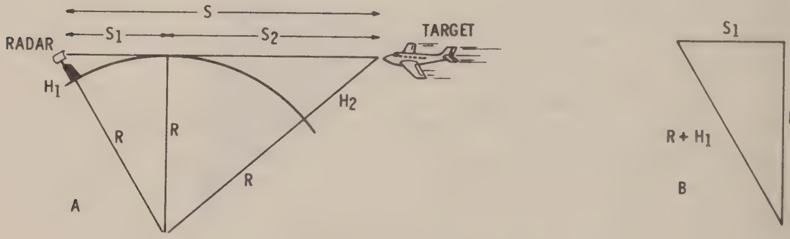


Fig. 2-19. Geometry of a distance problem.

$$(R + H_1)^2 = R^2 + S_1^2$$

$$R^2 + 2RH_1 + H_1^2 = R^2 + S_1^2$$

$$2RH_1 + H_1^2 = S_1^2$$

$$S_1^2 = H_1(2R + H_1)$$

Since H_1 (the antenna height) is negligibly small compared to 2R (twice the earth's radius), the second factor on the right $(2R + H_1)$ is nearly equal to 2R. Therefore, $(2R + H_1)$ may be replaced by the approximation 2R, so that $S_1^2 = 2RH_1$. This equation is satisfactory if S_1, R, and H_1 are in the same units. However, S_1 and R generally are expressed in miles, whereas H_1 is in feet. H_1 must be reduced to miles by the factor 1/5,280, since a foot is equal to 1/5,280 mile, and the equation becomes:

$$S_1^2 = \frac{2RH_1}{5280} = \frac{RH_1}{2640}$$

To determine the distance to the optical horizon, R is set equal to 4,000 miles and

$$S_1^2 = H_1 \left(\frac{4000}{2640}\right) = H_1 \left(\frac{500}{330}\right) = H_1 \left(\frac{50}{33}\right)$$

$$S_1^2 = H_1 \left(\frac{100}{66}\right)$$

The quantity, 100/66, is nearly equal to 3/2, or 1.5, so that by approximation $2_1^2 = 1.5H_1$.

To determine the distance to the *radar* horizon, R is replaced by 4/3 times 4,000 miles, and 1.5 must be replaced by 4/3 times 1.5, or 2.0. The equation then becomes, $S_1^2 = 2H_1$. For a point at height H above the water, then, the optical horizon is at a distance S_1 equal to $\sqrt{1.5H_1}$ or 1.23 $\sqrt{H_1}$, and the distance to the radar horizon is $\sqrt{2H_1}$, or 1.41 $\sqrt{H_1}$. Similarily, the radar distance, S_2, is $\sqrt{2H_2}$, or 1.41 $\sqrt{H_2}$. The range of radar, finally, is

$$S = S_1 + S_2 = \sqrt{2H_1} + \sqrt{2H_2}$$

Suppose, for example, that the radar is set at a height of 50 feet, and the target aircraft is at an altitude of 5,000 feet (Fig. 2-19A). The range from the radar to the horizon is S_1; this is calculated by taking the square root of twice the antenna height, or $\sqrt{2H_1}$. The range from the horizon to the target is S_2; this is calculated by taking the square root of twice the altitude of the target, or $\sqrt{2H_2}$, and where $2H_1$

equals 100 ft. and $2H_2$ is equal to 10,000 ft. The maximum possible range will be:

$$\sqrt{100} + \sqrt{10{,}000} = 10 + 100 = 110 \text{ miles}$$

Overwater Point-to-Point Communications Systems

In an overwater point-to-point communications system the reflected signal is important. It is necessary, therefore, in choosing antenna heights for a proposed installation to make sure that the direct and reflected signals do not cancel each other at the receiving point. It is possible, of course, to calculate the best antenna heights for use under standard conditions, but the computation is tedious and the results will not be correct for nonstandard conditions. Under the circumstances, antenna heights may be chosen by the rule-of-thumb method described here. Choose tentative antenna heights H_1 and H_2 which satisfy the relation:

$$\sqrt{2H_1} + \sqrt{2H_2} = S$$

where,

H_1 and H_2 are in feet,

and S is the distance between the proposed sites in miles.

Next, calculate a correction height, h, where,

$$h = 660 \sqrt{\frac{s}{f}}$$

where,

S is the same as before,

f is the frequency of operation in megacycles, and h is in feet.

The required height for each antenna is, then, the sum of the tentative height and the correction height, so that, by this rule, the antenna heights should be $(H_1 + h)$ and $(H_2 + h)$. If these values are inconvenient, choose new tentative heights and make a new computation.

OTHER LIMITATIONS ON MAXIMUM RANGE

In general, communications between two points is impossible when no path exists for a direct signal. This statement now can be phrased as follows: communications between two points is impossible unless an unobstructed radio line of sight exists. This does not mean, however, that communications is automatically possible whenever the line of sight is clear. Even with a clear line of sight, the signal level may be too low to be useful because of insufficient transmitter power, inadequate receiver sensitivity, or poor reflection from the target (in the case of radar). Besides this, there is always the possibility that the direct and reflected signals will arrive at the receiving point with nearly the same amplitude and with a phase difference of 180°. In this event, they will cancel each other, and communications will be impossible. Finally, various atmospheric effects may cause the curvature of the path followed by the signals to vary or may attenuate the signal severely so that communications is not possible.

MAXIMUM RANGE OF POINT-TO-POINT COMMUNICATIONS SYSTEMS

The maximum range over which a point-to-point communications system can reliably operate usually is limited by line-of-sight considerations. That is, the problems of long-range operation usually reduce to a question of finding satisfactory transmitting and receiving sites and building antenna towers that are sufficiently tall. With lightweight or portable equipment, however, it occasionally happens that the maximum reliable operating range is limited by considerations of available signal power. It is difficult to compute the maximum range over which a given system will operate reliably, but it can be shown that the maximum range depends on the characteristics of the system.

Power Density at Receiving Site

For any type of equipment the power actually delivered to the receiver is proportional to the power picked up by the receiving antenna. This, in turn, is proportional to the area of the receiving antenna and to the power density (power per unit area in a plane perpendicular to the direction of propagation) at the receiving site. The effectiveness of the transmitting equipment, therefore, is measured in terms of power density delivered at the receiving site, and this is proportional to P_tA/R^2, where P_t is the power output of the transmitter, A is the area of the transmitting antenna, and R is the distance from the transmitting site to the receiving site. P_t is entered in the expression as the first power in the numerator, since doubling the transmitter power will double the power density. A is equally logical. For example, if the transmitting-antenna width is doubled, the width of the radiated beam will be halved. Therefore the power in the beam will be distributed over one-half of the original area. This, of course, doubles the power density, and doubling the antenna width doubles the area of the antenna. The explanation of f is a little more complicated; doubling the frequency halves the wavelength and allows the antenna dimensions to be halved without altering the radiation pattern. If, therefore, the frequency is doubled, the antenna area may be cut to one-fourth of its original value (by halving both of its dimensions) without altering the power density. Following this, the antenna area may be quadrupled by restoring the antenna to its original size, and this will quadruple the power density. The power density is thus quadrupled by doubling the frequency, if the antenna area is unchanged. Finally, R is in the denominator as the second power, since doubling the range to the receiving point quadruples the area covered by the beam at the receiving point and cuts the power density to one-fourth.

Power Delivered to the Receiver

The power delivered to the receiver for a given power density is proportional to the area of the receiving antenna—since the larger the antenna, the greater the amount of power intercepted. Therefore since the power density delivered at the receiver site is proportional to P_tA/R^2, the power delivered to the receiver is proportional to $P_t\ Aaf^2/R^2$, where a is the area of the receiving antenna.

Maximum Operating Range

If the power delivered to the receiver is termed p, the expression above may be stated thus: p is proportional to P_tAaf^2/R^2.

The minimum usable receiver power, p', will be delivered to the receiver at the maximum reliable operating range, R', where, p' is proportional to P_tAaf^2/R'^2 and, by inversion, R' is proportional to f^2AaP/p'. From this, it follows that R' is proportional to $f\ \sqrt{AaP/p'}$.

On the basis of these equations the maximum reliable operating range theoretically can be doubled by doubling the frequency. Practically, however, this is not true because bad weather affects high-frequency signals more than it affects low frequency signals. This offsets the increase in maximum range to some extent, and even may make matters worse if the frequency is raised too far. In relation to other quantities the expression is more reliable. To double the maximum range by a change in a single quantity it is necessary to quadruple the area of one of the antennas, making either A or a four times as large as before; to quadruple the power output of the transmitter, making P four times as large as before; or to quadruple the power sensitivity of the receiver, making p' only one-fourth as large as before. The improvement, however, will be offset slightly by the fact that atmospheric effects will be more pronounced over the longer path when the range is increased.

MAXIMUM RANGE OF RADAR SYSTEM

The maximum operating range of a radar system is also normally limited by line-of-sight considerations. As with communications systems, however, lightweight or portable equipment may have its maximum range limited by considerations of signal power. The dependence of the maximum range on the system characteristics is easy to derive, since it follows directly from the expressions found in the preceding paragraph. The power density at the target is, as before, proportional to P_tA/R^2. This could be multiplied by the target area to find the power falling on the target. Instead, it is multiplied by a quantity called the *effective target area* in order to obtain the power in the portion of the reflected signal that travels back toward the radar. If the effective target area is b, the target acts like a transmitter with a power density that is proportional to P_tAbf^2/R^2. This expression then is divided by R^2 to obtain P_tAbf^2/R^4, which is proportional to the echo power density at the radar. Finally, it is multiplied by A, the area of the radar antenna, to obtain $P_tA^2bf^2/R^4$, which is proportional to the echo power picked up by the radar antenna. Then, as before,

$$p \text{ is proportional to } P_tA^2bf^2/R^4$$

$$p' \text{ is proportional to } P_tA^2bf^2/R'^4$$

$$R'^4 \text{ is proportional to } f^2A^2bP/p'$$

and the maximum range of the radar is proportional to $\sqrt[4]{f^2A^2\ bp/p'}$. It is important to note in this expression that the maximum range varies as the fourth root of P/p'. Because of this, a 10% change in the transmitter power output in the power requirement of the receiver will make only a 2.5% change in the maximum range on targets of a given type.

Chapter 3

Microwave Transmission Systems

When discussing the operation of radio and radar systems in the VHF region and at lower frequencies, it is customary to think of energy as being carried through space by electromagnetic waves and along transmission lines by voltages and currents. The voltage-current concept obviously is applicable in such conventional transmission systems as two-wire and coaxial lines, since both voltage and current can be measured readily. It is difficult, however, to apply the concept to a waveguide, which consists of a *single* hollow conductor on the outer surface of which neither voltage nor current can be detected. Although it is possible to understand the way a waveguide works in terms of voltages and currents, a clearer concept is gained by thinking of the energy as being carried by electromagnetic waves traveling inside the waveguide. The same approach, of course, can be applied to conventional transmission systems by considering the electromagnetic waves that accompany the voltages and currents on the lines.

In making use of the wave concept it is desirable to distinguish between guided and unguided waves. Guided waves are those that carry energy along transmission lines, and unguided waves are those that carry energy through space. The path of a guided wave is fixed by the guiding structure; that of an unguided wave is determined by the propagation characteristics of the region through which it travels. The signals radiated from an antenna are unguided waves, even though the antenna may be highly directional because the antenna controls only the direction in which the waves are launched —not the direction of subsequent travel.

TYPES OF TRANSMISSION LINES

Normal transmission systems used at microwave frequencies have either coaxial lines or hollow waveguides. A coaxial line is smaller and lighter than a waveguide but it has higher losses, except at S-band frequencies where the losses in short sections of coaxial line are not much greater than those in hollow waveguides. Normally, coaxial line is used for distances less than 15 feet, and waveguide for longer runs. At X- and K-band frequencies the losses of coaxial line are far greater than those of a waveguide, and, except for applications like coupling the local oscillator of a superheterodyne receiver to a waveguide transmission system, coaxial line seldom is used. The length of such a coaxial line normally does not exceed 1 or 2 inches.

Two-wire transmission lines (and one-wire lines in which the ground serves as the second conductor) radiate strong signals, unless the spacing between conductors is much less than a wavelength of the signal being carried. Such small spacing is difficult to achieve at microwave frequencies, however, and lines of this sort are not used. In addition to hollow waveguide, which guides an electromagnetic wave along its interior, a type of single-conductor transmission line called a surface waveguide, or G-line, sometimes is used at microwave frequencies. A G-line guides a wave that travels through the region surrounding the one-wire line, and, like a waveguide, it requires no second conductor.

The coaxial line has only two important applications at microwave frequencies. It is used as a transmission line in the S-band over distances of approximately 15-feet or less, and in the X- and K-bands over distances of approximately 2 inches or less. Its merit for S-band operations lies in the fact that it is lighter and less bulky than waveguide and, although its losses are greater than those of a waveguide, the difference in loss is not significant for short runs. For runs over 15 feet, waveguide generally is used. In the K- and X-bands, coaxial line is used to couple the output of a low-powered oscillator to a waveguide transmission line, because it is fabricated and more easily handled than waveguide. However, in these bands coaxial-line losses are so much greater than those of hollow waveguide that it is never used.

S-Band Coaxial Line

To keep losses as low as possible, S-band coaxial line uses air as an insulator between the inner and outer conductors. The line usually is filled with dry air, and the pressure inside is held at approximately 5 pounds per square inch above atmospheric pressure so that any leak results in the loss of dry air rather than an invasion of outside air, which may be moist. The use of conventional insulating material (other than air) is avoided by employing short sections of coaxial line in order to support the center conductor of the main line (Fig. 3-1). If a transmission line is exactly a quarter wavelength and is terminated in a short circuit, its input impedence is extremely high. The supports, called quarter-wave stub supports, are sections of this type. Their impedance at the operating frequency is so high that they do not affect the characteristics of the line, although, at most other frequencies, they look like low-impedance shunts across the line (for direct current they look like a short circuit). Because of this type of construction, coaxial line is rigid, and joints must be provided wherever one section of a line moves while an adjacent section is stationary.

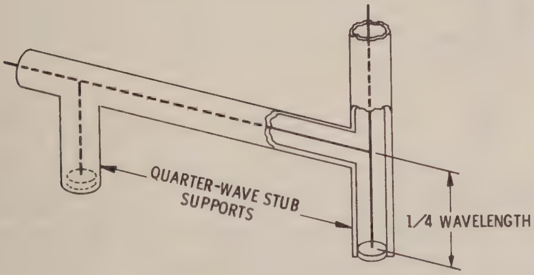


Fig. 3-1. Coaxial line with quarter-wave stub supports.

X- and K-Band Coaxial Line

A typical use of coaxial line in the X- and K-bands is illustrated in Fig. 3-2, which shows an oscillator tube for X-band operation. The tube has a conventional octal base, and the coaxial line replaces one of the pins. The corresponding location at the socket is left clear so that the line can project through the socket and into the waveguide, as shown in Fig.

3-3. The coaxial line thus serves as a coupling between the tube and the transmission line. Here, the coaxial line consists of silver inner and outer conductors with a solid insulator of polystyrene. The outer conductor is terminated at a point even with the upper surface of the waveguide, and the inner con-

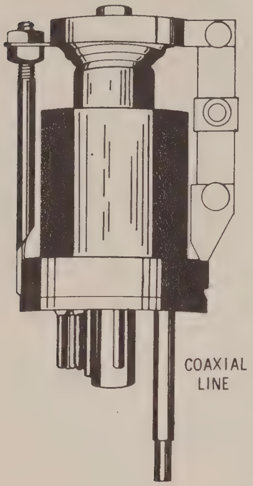

COAXIAL
LINE

Fig. 3-2. Oscillator tube with coaxial line connected.

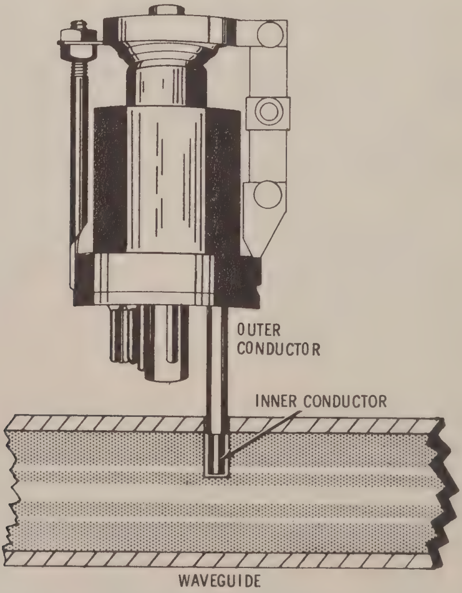

OUTER
CONDUCTOR
INNER CONDUCTOR
WAVEGUIDE

Fig. 3-3. Oscillator tube coupled to waveguide.

ductor is extended into the guide. The polystyrene insulator also is extended and serves to protect the rather fragile inner conductor. The extended inner conductor acts as an antenna and radiates energy into the space inside the waveguide. The losses in such a coaxial line are relatively high, but over such a short distance, they are unimportant.

ELECTRIC AND MAGNETIC FIELDS IN COAXIAL LINE

Although the behavior of a conventional transmission line is explained adequately in terms of voltages and currents, the electric and magnetic fields associated with these should be examined in order to establish a background for the explanation of waveguide behavior. In the following discussion it is assumed that the line is a coaxial line having no losses.

Electric Field

An electric field is directed from the positive to the negative conductor at any point in the space between the coaxial-line conductors. At a given point, the strength of the field is proportional to the voltage difference between the conductors. For a given voltage, the strength of the field is inversely proportional to the distance between the axis of the line and the point in question. In Fig. 3-4 the arrows show the direction of the electric field when the outer conductor is positive and the inner conductor is negative. The field strength falls off as the distance from the center of the inner conductor increases.

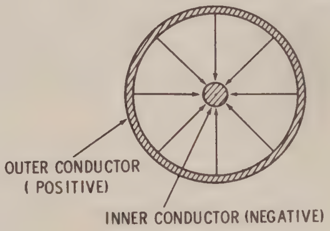


Fig. 3-4. Electric field in coaxial line.

Magnetic Field

The direction of the magnetic field in a coaxial line is always tangent to a circle the center of which is on the axis of the line, as shown in Fig. 3-5. At a given point, the strength of a magnetic field is proportional to the current flow. For a given current, the strength of the magnetic field varies inversely with distance from the axis of the line (just as the strength of the electric field varies). A reversal in current direction causes a reversal in the direction of the magnetic field.

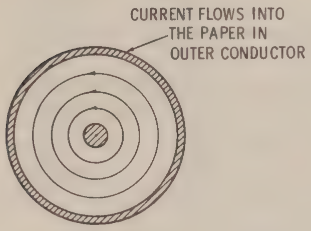

CURRENT FLOWS INTO THE PAPER IN OUTER CONDUCTOR

Fig. 3-5. Magnetic field in coaxial line.

Combined Fields

Fig. 3-6 shows the combined electric and magnetic fields in a coaxial line that is carrying a signal in only one direction. The line is shown cut in half along its axis, and the plane along which it is cut has been shaded to show the strength of the fields. The signs and arrows above the line show the sign of the voltage on the outer conductor and the direction of current flow in the outer conductor. The straight arrows to the right of the shaded cut section show the direction of the electric field either toward or away from the center conductor. The curved arrows to the right show the direction of the magnetic field, either clockwise or counterclockwise around the inner conductor. The line is assumed to be lossless so that

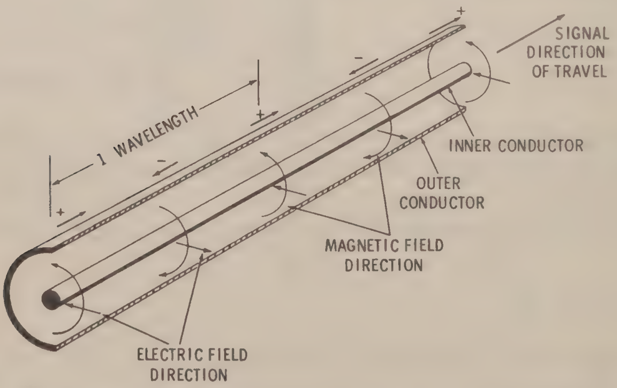
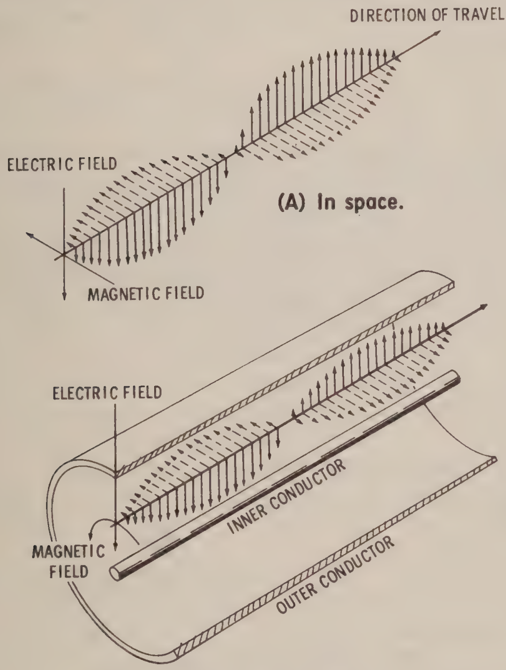

+
−
+
−
+
1 WAVELENGTH
SIGNAL DIRECTION OF TRAVEL
INNER CONDUCTOR
OUTER CONDUCTOR
MAGNETIC FIELD DIRECTION
ELECTRIC FIELD DIRECTION

Fig. 3-6. Combined electric and magnetic fields in coaxial line.

characteristics impedance Z_o is a pure resistance. Voltage and current are then in phase, and peak voltage and peak current occur at the same time at a given point. Similarly, both voltages and current vanish simultaneously at a given point. Since E equals IR at all points on the line (where R is the resistance that is equal to Z_o), the voltage is always proportional to the current, and the electric field is always proportional to the magnetic field. The same shading then serves to show the strength of both fields.



(A) In space.

(B) In coaxial line.

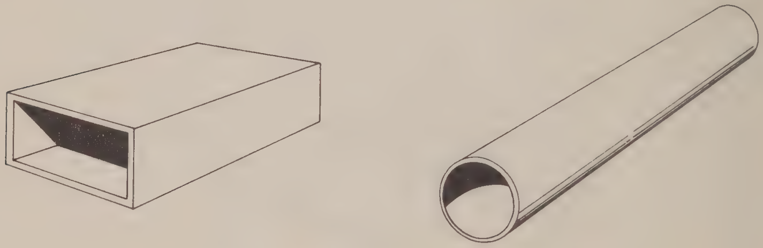
Fig. 3-7. Electromagnetic waves in space and coaxial line.

Electromagnetic Wave

Fig. 3-7A shows the strength and direction of the electric and magnetic fields for points lying on a line parallel to the path of an electromagnetic wave traveling through space. The direction of the electric or magnetic field at any point is shown by the direction of the arrow, the tail of which lies at the point in question. (One arrow shows the electric field; the other shows the magnetic field.) The strength of the field is shown by the length of the arrow. In the same way, Fig. 3-7B shows the electric and magnetic fields for points on a line parallel to the axis of a coaxial transmission line. From the similarity in the two field arrangements, it can be seen that the fields in the coaxial line form an electromagnetic wave which travels through space between the inner and outer conductors in exactly the same way that an electromagnetic wave travels through unobstructed space. It is, therefore, easy to think of energy as being carried down the coaxial line by an electromagnetic wave (although, of course, it is even easier to think of it as being carried by voltage and current). In the following section the electromagnetic wave concept is used to explain how energy is carried through the space inside a hollow waveguide.

WAVEGUIDES

Although any transmission system serves as a guiding structure for guided waves, the term *waveguide* is reserved for the type that consists of a single hollow conductor. Although a waveguide may have almost any shape, the two types in common use are those having rectangular and circular cross sections (Fig. 3-8). The characteristics of the rectangular waveguide shown in Fig. 3-8A are determined by the height and width. (These dimensions are always the inside measurements, the smaller of the two customarily being taken as the height, even though it may lie in a horizontal plane.) The characteristics of the circular waveguide shown in Fig 3-8B are determined by the diameter, which is always the inside measurement.



(A) Rectangular. (B) Circular.

Fig. 3-8. Typical waveguide types.

Transmission in a Rectangular Waveguide

Although a rectangular waveguide can carry energy in many ways, only one way, the simplest, is practical. In this case, vertically polarized waves travel down the waveguide along paths that are parallel to the top and bottom of the guide. As stated, the electrical characteristics of a rectangular waveguide are determined by the inside dimensions (Fig. 3-9).

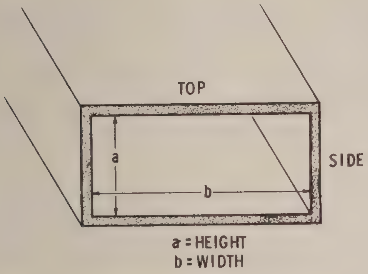
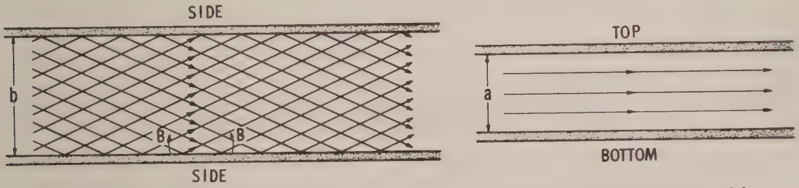


Fig. 3-9. Rectangular waveguide dimensions.

Whatever the orientation of the waveguide, it is customary to call the shorter dimension the height, and the longer one the width. Following this thought, the longer walls of the waveguide are called the top and bottom, and the shorter walls are called the sides. Vertical polarization as applied to an electromagnetic wave in a rectangular waveguide, therefore, means that the electric field is parallel to the shorter dimension of the guide. The small letters *a* and *b* customarily are used to denote the height and width, respectively, when referring to rectangular waveguide dimensions.

The paths followed by electromagnetic waves in the waveguide are shown in Fig. 3-10. Fig. 3-10A is a side view, with the wave paths parallel to the top and bottom walls of the waveguide. Fig. 3-10B is a top view of the paths. All of the paths make the same angle, *B*, with the sides of the waveguide.



(A) Top view of waveguide.
(B) Side view of waveguide.

Fig. 3-10. Wave paths in rectangular waveguide.

Determination of Angle B—Angle B cannot be chosen arbitrarily. Instead, it depends on *b*, the width of the waveguide, and on λ, the wavelength of the signals. The relationship is given by the equation,

$$\sin B = \frac{\lambda}{2b}$$

This equation results from the fact that propagation is impossible unless all signals traveling in a given direction reinforce each other. The equation is derived easily from consideration of the diagram shown in Fig. 3-11. Consider a signal that passes through *C* moving to the left. It is reflected from the left-hand wall at *D*, from the right-hand wall at *E*, and finally moves in direction EF, which is parallel to the original direction, CD.

A wavefront is always perpendicular to the path of a wave; and since DH is perpendicular to CD, DH is a wave front for signals moving to the left. As explained in the first part of Chapter 2, the phase of a signal is the same at all points on a wavefront. A signal moving to the left through *H* is, therefore, in phase with the incident signal at *D*. If energy is to be propagated down the waveguide, the reflected signal at *E* must have the same phase as the signal moving to the left through *I*, since both are on a wave front perpendicular to EF. Otherwise, the signals will cancel each other after traveling a short distance through the waveguide. Since the electric field of the signals is parallel to the reflecting wall, there is one phase reversal at *D* and another at *E*. The signal behavior on reflection is the same as that of a horizontally polarized signal reflected from the surface of the sea (see Chapter 2). However, it does not matter whether this is the case, since two phase reversals have a net effect of zero. To have the desired phase relations, angle B must have a value which makes path

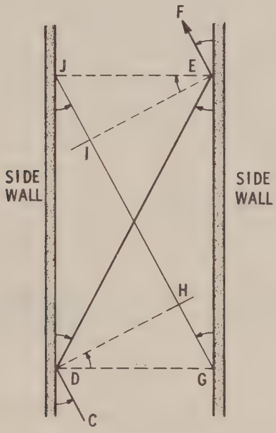


Fig. 3-11. Equation geometry for determination of angle.

DE an integral (whole) number of wavelengths longer than path *HI*. GJ is equal to DE, since *G* lies opposite *D*, and *J* lies opposite *E*. The requirement for propagation of energy is then that HI and GJ differ by an integral number of wavelengths. The difference between GJ and HI is the sum of GH and IJ, and each of these distances is equal to *b* sin *B*. The condition for propagation, then, is that

$$2b \sin B = n\lambda$$

where,

n is an integer 1, 2, 3, and so on.

In the simplest case, which is the one under discussion, *n* is equal to 1 and the preceding equation immediately reduces to

$$\sin B = \frac{\lambda}{2b}$$

when 1 is substituted for *n*.

It must be emphasized that the waveguide will carry energy only for certain values of angle *B*, because the signals cancel for other values of the angle. This has nothing to do with attenuation or losses; it is similar to the directive effect of an antenna that does not radiate in directions for which signals from various portions of the antenna cancel each other.

Field Patterns in a Rectangular Waveguide

The patterns of the electric and the magnetic fields in the rectangular waveguide may be found by adding the fields belonging to the two sets of electromagnetic waves—those mov-

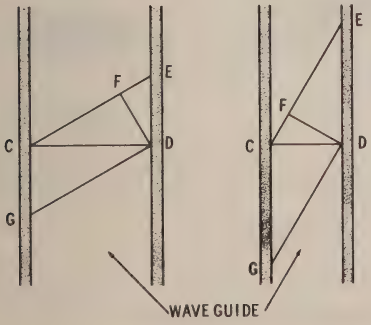
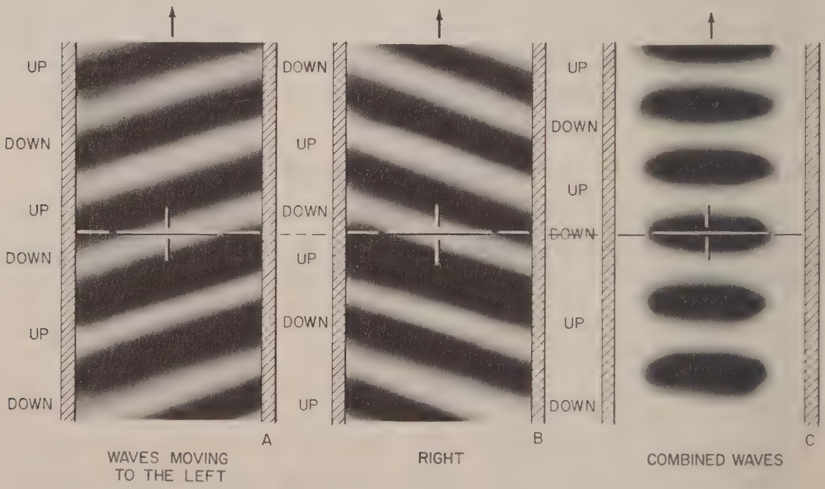


Fig. 3-12. Spacing of wavefronts in the waveguide.

ing to the left and those moving to the right. This may be done separately for the electric and the magnetic fields. In order to construct the field patterns, it is necessary to know the spacing in the waveguide between two wave fronts having a phase difference of one-half cycle. This is developed in the diagrams shown in Fig. 3-12. Here, CD is a line drawn across the waveguide, and CE is a wave front of the signals that is moving to the left. The preceding discussion has shown that DF (which corresponds to GH in Fig. 3-11) must be one-half wavelength in the simple case (for which *n* is 1). Wave front DG in Fig. 3-12 is therefore one-half cycle different in phase from wave front CE.

Pattern of Electric Fields

Fig. 3-13 illustrates the way in which the electric fields combine. Fig. 3-13A shows the electric fields of the waves that are moving to the left, and Fig. 3-13B shows those of the waves moving to the right. In each illustration the direction of the electric field is indicated by the notes beside the figure, and the strength of the field is shown by the darkness of the shading. Fig. 3-13C shows the combined fields that are obtained simply by adding the fields of the two sets of waves.



(A) Waves moving to the left.

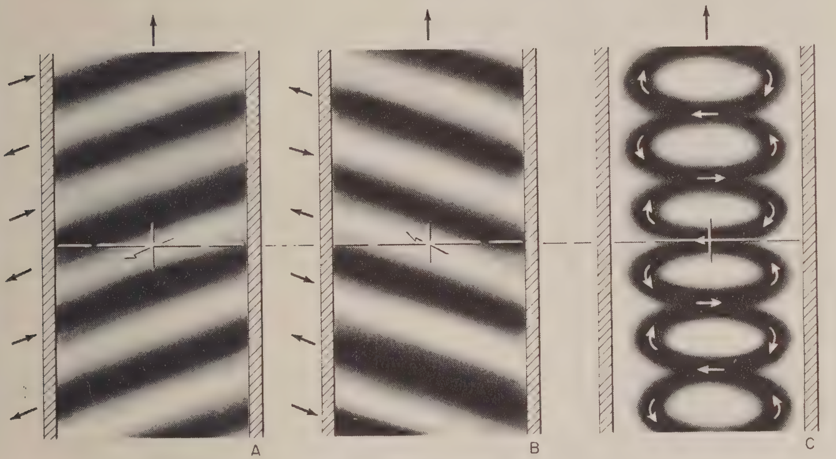
(B) Waves moving to the right.

(C) Combined waves.

Fig. 3-13. Electric-field patterns.

Pattern of Magnetic Fields

Fig. 3-14 illustrates the way in which the magnetic fields combine. When adding these fields, it is necessary to consider the relative directions of the individual fields, in the same way that, when adding two AC voltages, it is necessary to consider the relative phases. The direction of the magnetic field is shown in Figs. 3-14A and B by small arrows next to the waveguide, and in Fig. 3-14C, by arrows superimposed on the shading. The strength of the fields is indicated by the darkness of the shading.



(A) Waves moving to the left.

(B) Waves moving to the right.

(C) Combined waves.

Fig. 3-14. Magnetic-field patterns.

Combined Pattern of Electric and Magnetic Fields

The combined pattern of the electric and magnetic fields is shown in Fig. 3-15. Here, the strength of the electric field is indicated by the darkness of the shading; the direction is given as a note beside the waveguide. The character of the magnetic field is indicated by the closed loops. The arrowheads on these loops show the direction of the magnetic field, and the loop spacing indicates the magnetic-field strength; closely spaced loops indicate a strong field. The magnetic field pattern is shown more precisely in Fig. 3-14C. At both side walls the electric field vanishes and the magnetic field is longitudinal (parallel to the axis of the waveguide). Midway between the

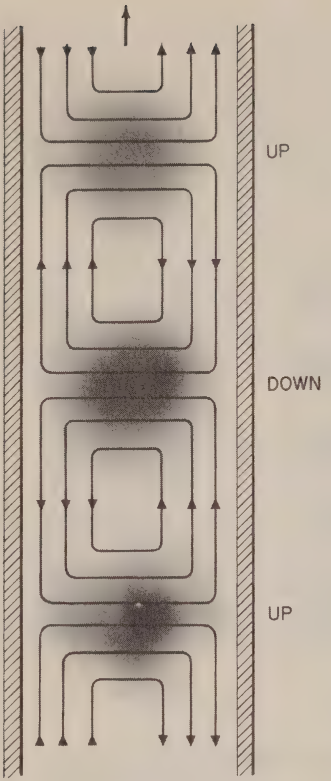
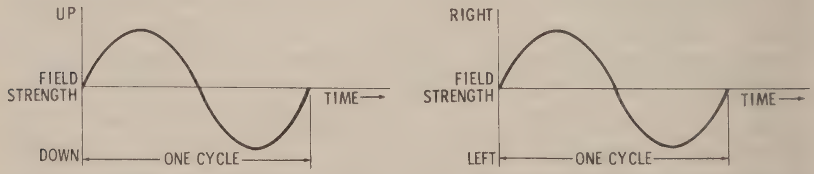


Fig. 3-15. Electric and magnetic fields combined patterns.

side walls the field is the same as in unobstructed space, both electric and magnetic fields being transverse (perpendicular to the waveguide axis). At intermediate points, fields of both sorts are present with reduced strength, the longitudinal and traverse magnetic fields adding to produce a combined magnetic field that is neither transverse nor longitudinal.

At any point inside the waveguide sinusoidally oscillating fields will be observed as the field pattern travels down the guide past the point of observation. The character of the observed fields depends only on the distance between the point

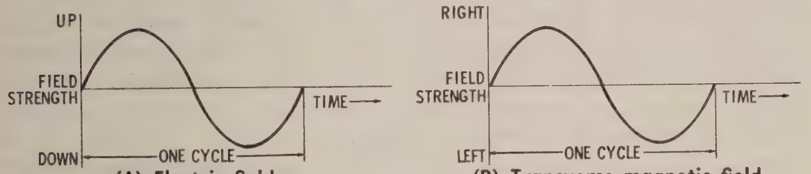


(A) Electric field.

(B) Transverse magnetic field.

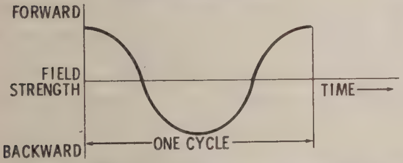
Fig. 3-16. Electric and magnetic fields as observed between the side walls.

of observation and the waveguide side wall. At a point midway between the side walls of the waveguide the electric field (Fig. 3-16A) and the transverse magnetic field (Fig. 3-16B) vary. The two fields are in phase with each other, the peak intensity of the electric field occurring at the same instant as the peak intensity of the transverse magnetic field. No longitudinal magnetic field is observed at a point midway between the side walls. At a point somewhere between the center of the guide and one of the side walls the fields vary, as in Figs. 3-17A and B. These figures show the fields for a point between the center of the guide and the right-hand side wall. For a point between the center of the guide and the left-hand wall, the observed fields are the same as those in Fig. 3-17, except that the phase of the longitudinal magnetic field in Fig. 3-17C is reversed. At both side walls, the longitudinal magnetic field is present but both the transverse magnetic field and the electric field vanish.



(A) Electric field.

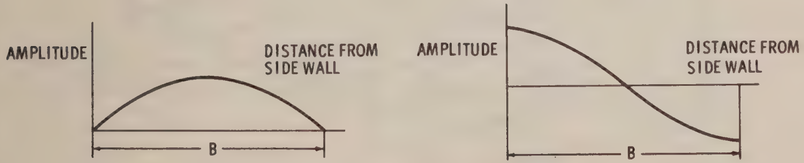
(B) Transverse magnetic field.



(C) Longitudinal magnetic field.

Fig. 3-17. Fields observed at a point between the center of the guide and the right-hand wall.

If the amplitude of the oscillating field is plotted against distance from the side wall of the waveguide, the resulting curve is half of a sine wave (Fig. 3-18). For this reason the pattern sometimes is called a *half-sine* pattern. The electric



(A) Electric and transverse magnetic fields.

(B) Longitudinal magnetic field.

Fig. 3-18. Field amplitudes.

and transverse magnetic fields vanish at both walls, as in Fig. 3-18A. The longitudinal magnetic fields vanishes at the center of the guide, as in Fig. 3-18B, and it has opposite phases in the right and left halves of the guide, as shown by the negative value of the amplitude in the figure.

Cut-Off Wavelength

As explained in Fig. 3-11, angle *B* is determined by the relationship:

$$\sin B = \frac{\lambda}{2b}$$

for the simplest wave pattern that will carry energy down the waveguide. This equation shows that *B* must equal 90° when the wavelength is equal to 2b. Here, the two sets of waves merely travel back and forth across the waveguide and no energy is carried in the desired direction (along the guide). If the wavelength is greater than 2b, there is no value of *B* that satisfies the relation and a wave pattern of the sort described in the previous discussion cannot exist. The wavelength at which *B* equals 90° is called the λ_c (cut-off wavelength), and the corresponding frequency is called the f_c (cut-off frequency). For the wave pattern described above, λ_c is equal to 2b. To avoid confusion the wavelength of the signal is referred to as the λ_a (wavelength in air). (A signal with a

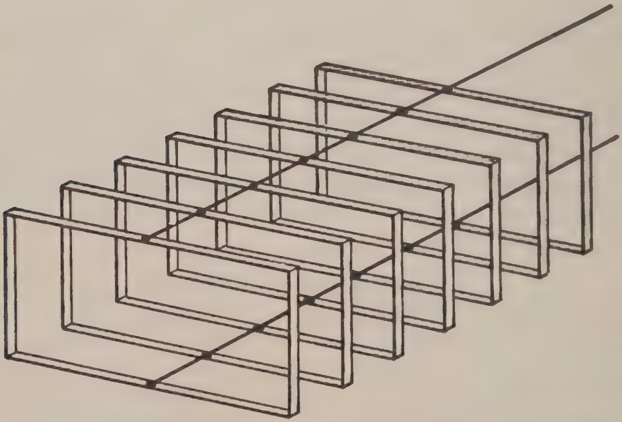


Fig. 3-19. Two-wire transmission line with quarter-wave shorted stubs.

frequeny of 3,000 mc, for example, has λ_a equal to 10 centimeters.) Using this terminology a waveguide cannot transmit energy by means of the simple pattern *unless* λ_a is shorter than λ_c.

Waveguide transmission sometimes is explained in terms of a two-wire transmission line with an infinite number of quarter-wave shorted stub, each one of which has an infinite impedance and therefore looks like an open circuit. Fig. 3-19 shows a two-wire line with a large number of stubs, and the similarity to a waveguide is immediately apparent. This explanation does not cover the field pattern nor the group and phase velocities, which are discussed subsequently. It further implies that transmission is possible *only* if the width of the guide is one-half wavelength, which is incorrect since transmission is possible only when the width of the guide *exceeds* one-half wavelength. It may be useful to remember, however, that a waveguide at cutoff frequency (where it will not carry energy) looks like a two-wire line with an infinite number of quarter-wave shorted stubs.

Guide Wavelength

In addition to λ_a and λ_c, it is customary to refer to the λ_g (guide wavelength). This is the wavelength of the pattern in the waveguide, and it is equal to $\lambda_a/\cos B$. The relationship is derived easily from the diagram in Fig. 3-20. Here, *CD* is a wave front of the signal that moves to the left. *E* is on the left-hand wall directly opposite *D*, and *EF* is another wave front of the same signal. Distance *DG* is equal to *b* sin *B* and, from the discussion of Fig. 3-11, it is clear that this must be $\lambda_a/2$. The two wave fronts, *CD* and *EF*, are therefore separated by one-half λ_a. *H* is on the left-hand wall directly op-

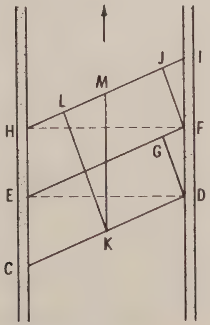


Fig. 3-20. Equation geometry for derivation of the guide wavelength.

posite *F*, and *HI* is another wave front. By an argument similar to the one given above, the spacing between *EF* and *HI* is one-half λ_a. Then *KL*, the spacing between CD + HI is exactly equal to λ_a. The wavelength of the pattern, however, is equal to *KM*—since this is the distance along the guide axis between two points where the phase of the signal differs by 1 cycle. The angle between *KL* and *KM* is B, therefore:

$$\mathrm{KD} = \lambda_a = \mathrm{KM} \cos \mathrm{B} = \lambda_g \cos \mathrm{B}$$

and

$$\lambda_g = \frac{\lambda_a}{\cos \mathrm{B}}$$

GROUP AND PHASE VELOCITIES

Just as two wavelengths, λ_a and λ_g, must be considered in connection with the transmission of energy along a waveguide, three separate velocities are of concern. The first of these is the velocity of electromagnetic waves in space, usually written *c*, which is equal to 330 yards, or 300 meters, per microsecond. The other two are the v_g (group velocity), the velocity with which energy is transmitted along the waveguide, and the v_p (phase velocity), the velocity with which the field pattern moves along the waveguide.

Distinction Between Velocities

Imagine a transmission system, similar to the one shown in Fig. 3-21, that consists of a vertical glass tube through which

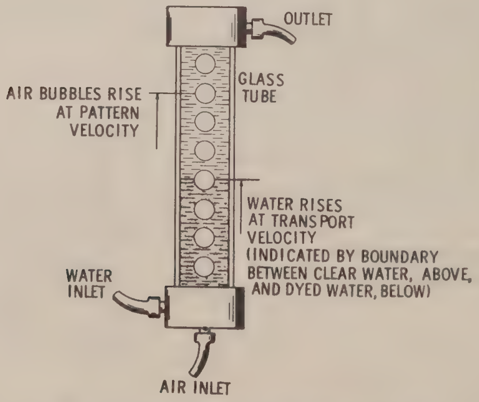

OUTLET
GLASS TUBE
AIR BUBBLES RISE AT PATTERN VELOCITY
WATER RISES AT TRANSPORT VELOCITY (INDICATED BY BOUNDARY BETWEEN CLEAR WATER, ABOVE, AND DYED WATER, BELOW)
WATER INLET
AIR INLET

Fig. 3-21. Pattern and transport velocities.

water is pumped. Simultaneously, air is introduced at the bottom and rises in the form of small, regularly spaced bubbles. Suppose now that clear water is pumped through the tube. The rising bubbles will form a pattern which moves upward at a speed that may be called the *pattern velocity*. If dyed water then is introduced, the boundary between the clear and colored water will rise at a speed that may be called the *transport velocity*—the velocity at which water moves through the tube. The case is similar to that of a waveguide. If a steady signal is sent through the guide, the field pattern will have a wavelength, λ_g, analogous to the spacing between the bubbles. The field pattern will move with a phase velocity, just as the bubble pattern moves at the pattern velocity. If, then, the signal level is suddenly raised, the boundary between the low- and high-level signals will move along the waveguide with a group velocity, just as the boundary between the clear water and the dye solution moves at the transport velocity.

Variation of Velocities With Wavelength

As explained, the cut-off wavelength λ_c, is equal to 2b, and angle B is determined by the relation:

$$\sin B = \frac{\lambda_a}{2b} = \frac{\lambda_a}{\lambda_c}$$

When the operating frequency is extremely high so that the wavelength of the signal (in air) is very much shorter than the cut-off wavelength of a waveguide, sin *B* and *B* are both very small and cos *B* is nearly 1. In this case, the phase velocity is only slightly greater than *c* and the group velocity is only slightly smaller. At the other extreme, when the wavelength in air is only slightly shorter than the cut-off wave-

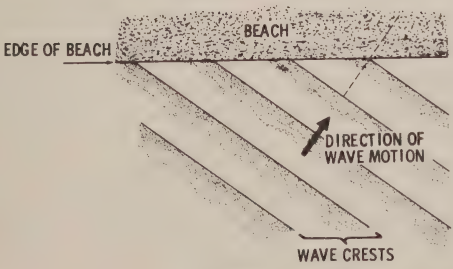


Fig. 3-22. Water waves breaking on a stretch of beach.

length, sin B is nearly 1 and B is nearly 90°. Cos B is then very small, and v_g becomes small and v_p becomes extremely large. Similar behavior can be observed at any straight section of beach. Fig. 3-22 shows a series of wave crests moving in a direction that makes an angle, X, with the edge of the beach. As a point on one of the wave crests moves from P to Q, the intersection of that crest with the beach moves from R to Q. Using simple geometry, distance PQ is equal to distance RQ times cos X. Then, if the waves move forward with velocity v_w and the point at which the wave crest breaks on the beach moves along the edge of the beach with a velocity v_b,

$$PQ = v_w t, \quad RQ = v_b t$$

where t is the time required for the wave crest to move from P to Q, and

$$v_w t = v_b t \cos X$$

so that

$$v_b = v_w / \cos X$$

The point of intersection of a wave crest with the edge of the beach therefore runs along the beach 1/cos X times as fast as the wave pattern moves.

OTHER MODES OF TRANSMISSION

A number of individual electromagnetic waves can combine in an infinite number of ways to carry energy down the interior of a rectangular waveguide. The various ways are called *modes*, and each mode has a distinctive field pattern that differs from that of every other mode. Each mode also has its own cut-off frequency, cut-off wavelength, guide wavelength, group velocity, and phase velocity—not necessarily different from those of other modes. In a rectangular waveguide it is customary to choose dimensions that make the cut-off frequency higher than the operating frequency for all modes except the simple one used as an example in the previous discussion. For all modes except this one the operating frequency is below the cut-off frequency (and the operating wavelength is longer than the cut-off wavelength), making transmission impossible. The waveguide, therefore, operates only in the simple mode. In spite of this practice, the other modes should be considered briefly.

Mode Terminology

In some modes the electric field is always transverse (where it does not vanish), whereas the magnetic field is sometimes transverse and sometimes longitudinal. This is true, for example, with the simple mode described previously. Such modes are called TE (transverse electric) modes. In other modes the magnetic field is always transverse, whereas the electric field is sometimes transverse and sometimes longitudinal. These are called TM (transverse magnetic) modes. In still other modes both the electric and magnetic fields are longitudinal at various points. These modes are not named, since a mode of this sort always can be considered as a combination of a given TE mode with a given TM mode.

A mode is defined further by two subscripts, *m* and *n*, as a $\mathrm{TE}_{m,n}$ or a $\mathrm{TM}_{m,n}$ mode. The first subscript is the number of times a pattern repeats itself along the *a* dimension of the waveguide (from the top of the guide to the bottom) and the second subscript is the number of times that the pattern repeats itself along the *b* dimension (from one side of the waveguide to the other). In the simple case described earlier, the field pattern is uniform from top to bottom, so that there is *no* repetition, and the first subscript is therefore *O*. There is variation, however, from one side to the other; but the pattern occurs only once, and therefore, the second subscript is *1*. The simple mode is, accordingly called the $\mathrm{TE}_{0,1}$ mode (since the electric field is entirely transverse). The following discussion shows that the $\mathrm{TE}_{0,1}$ mode has a longer cut-off wavelength than any other mode for a given rectangular waveguide. Because of this, it is possible to choose the waveguide dimensions so that λ_c is longer than λ_a for this mode, and shorter than λ_a for every other mode. The waveguide is thereby forced to operate in the $\mathrm{TE}_{0,1}$ mode, as explained.

Two Nonexistent Modes

It is impossible to send a single vertically polarized wave down a waveguide along a path parallel to the axis of the guide. The field pattern of such a wave will have an electric field that is parallel to each side wall at the surface of the wall. This will cause currents to flow in the side walls and the electromagnetic waves radiated by these currents will cancel the original signal in a very short distance. The same thing is, of course, true of a horizontally polarized wave that travels directly down the guide without reflection. In this case, there will be electric fields parallel to the top and bottom of the guide at the top and bottom, and the same result will be obtained. Finally, any electromagnetic wave may be considered as a combination of vertically and horizontally polarized waves. It follows, therefore, that it is impossible to send energy through a rectangular waveguide by means of any signal that travels directly down the guide without reflection. Now, the $TE_{0,0}$ and $TM_{0,0}$ modes have field patterns that are uniform along both the a and b dimensions. It was explained earlier that reflection from the side walls causes the field pattern to vary along the b dimension, whereas the absence of reflection from the top and bottom of the waveguide makes the pattern uniform along the a dimensions. Thus, a field pattern uniform along the a and b dimensions can be obtained only if there is no reflection from any of the walls. It is, however, impossible to transmit energy through a rectangular waveguide by means of waves that are not reflected; and for this reason, neither the $TE_{0,0}$ nor the $TM_{0,0}$ mode can exist in a rectangular guide.

$TE_{0,n}$ Modes

It has been stated that energy can be carried along a waveguide by vertically polarized waves traveling parallel to the top and bottom of the guide, if their paths make the proper angle with the side walls. Also, the proper angle is given by the equation,

$$2b \sin (B) = n\lambda_a$$

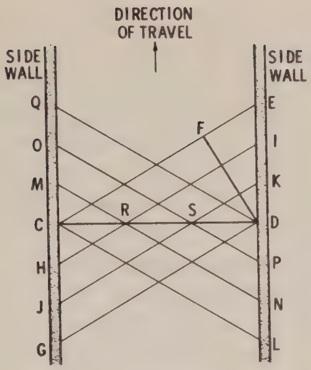
or by

$$\sin (B) = \frac{n2_a}{2b}$$

where B is the angle between the paths of the individual waves and the side walls.

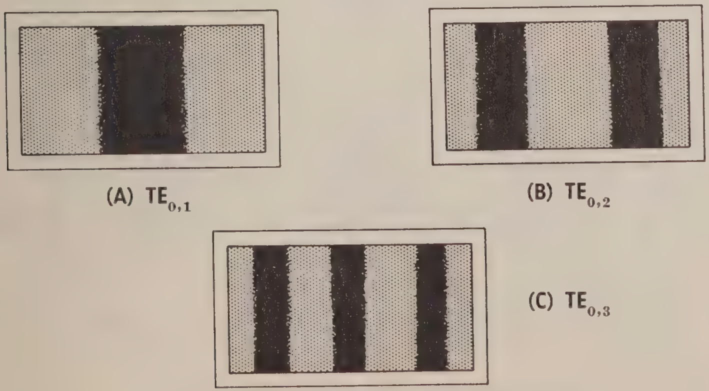
It has been shown that the $TE_{0,1}$ mode is produced when n is equal to 1, although the mode was not so labeled at the time. For any value of n (except zero), this arrangement of vertically polarized waves reflected from the side walls produces the $TE_{0,n}$ mode. This may be verified by examination of the diagram shown in Fig. 3-23, which treats the case of n equal to 3. Here, C and D are two points opposite each

Fig. 3-23. Geometry of the $TE_{0,3}$ mode.



other on the side walls of a rectangular waveguide. Distance DF is *n* half-wavelengths (3 in this case) and wave front GD is, therefore, 3 half-wavelengths behind wave front CE. It is, then, possible to construct additional wavefronts HI and JK so that each is one-half wavelength from its neighbors. In the same way, wavefronts QD, OP, MN, and CL may be constructed, each one-half wavelength from its neighbors.

Now, as one moves from *C* to *R*, the phase of the waves moving to the left changes one-half cycle in one direction, and the phase of the waves moving to the right changes an equal amount in the other direction. This leads to a pattern that is similar to that obtained by moving *all the way* across the guide in the $TE_{0,1}$ mode (Figs. 3-12 through 3-15). Conditions at *R* are the same as those at *C*, except that the phases of both sets of waves are reversed. The pattern between *R* and *S* is, therefore, the same as the pattern between *C* and *R*, but the directions of the fields are reversed. The same thing


(A) $TE_{0,1}$

(B) $TE_{0,2}$

(C) $TE_{0,3}$

Fig. 3-24. Electric field patterns for $TE_{0,n}$ modes.

happens between *S* and *D*, with the directions of the fields the same as those between *C* and *R*. Consequently, the pattern for a $TE_{0,n}$ mode is nothing more than a repetition of the $TE_{0,1}$ pattern with the field directions reversed at each repetition. Fig. 3-24 illustrates the electric field in a plane perpendicular to the axis of a rectangular waveguide; Fig. 3-24A shows the $TE_{0,1}$ mode, Fig. 3-24B shows the $TE_{0,2}$ mode, and Fig. 3-24C shows the $TE_{0,3}$ mode. In Fig. 3-24B the direction of the electric field in the right half of the waveguide is opposite to that in the left half. In one half, the field is directed upward; in the other it is directed downward. Similarly, in Fig. 3-24C the direction in the center of the guide is opposite to the direction in either of the side portions of the guide. In every case, however, the field is either upward or downward. Fig. 3-25, like Fig. 3-15, shows the combined fields for the $TE_{0,3}$ mode.

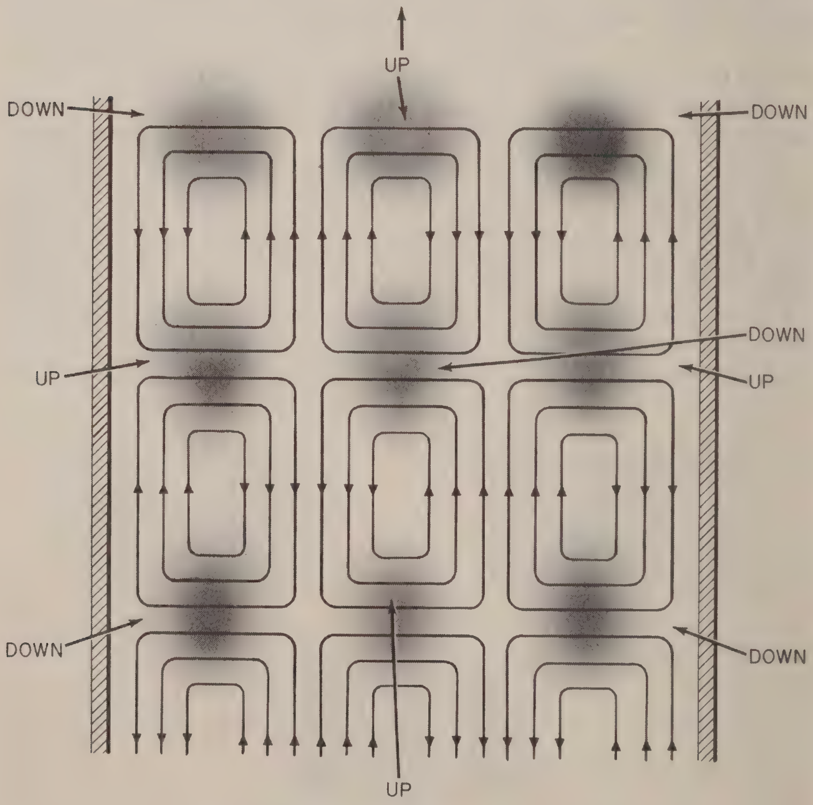


Fig. 3-25. Combined fields for the $TE_{0,3}$ modes.

Other Nonexistent Modes

At first glance it would appear possible to develop the $TM_{0,n}$ mode by sending horizontally polarized signals along the paths prescribed for the signals that produce the $TE_{0,n}$ mode. This will produce a pattern of the desired type, but there will be longitudinal electric fields at the walls and the pattern will be cancelled by the waves radiated by induced wall currents. The same objection applies to the possibility of developing a $TM_{m,0}$ mode by sending vertically polarized waves along the paths that produce a $TE_{m,0}$ mode. For this reason, neither the $TM_{0,n}$ modes nor the $TM_{m,0}$ modes can exist in a rectangular waveguide.

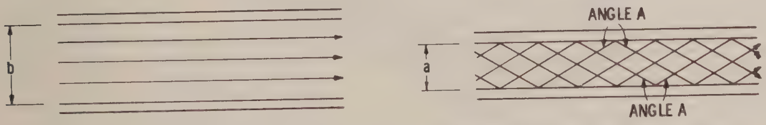
Reflection From Four Walls

All of the modes previously considered have patterns that are built up from two sets of waves. Before considering the $TE_{m,n}$ and $TM_{m,n}$ modes, it is necessary to consider what happens when the signals are reflected from all four walls of the waveguide. The pattern then will be built up from four sets of waves—those moving downward to the right and left and those moving upward to the right and left. To begin, consider the two sets of waves that move downward to the right and left. Their paths meet the side walls of the waveguide at some angle, *B*, which satisfies the equation

$$\sin B = \frac{n\lambda_a}{2b}$$

and the two sets of waves combine to form a field pattern that is similar to that of the $TE_{0,n}$ mode, except that it is tipped downward and travels downward.

A similar pattern, tipped in the opposite direction and traveling upward, is formed by the two sets of waves that move upward to the right and left. These two patterns follow paths that meet the top and bottom of the waveguide at an angle, *A*. Fig. 3-26A shows the paths from above and Fig. 3-

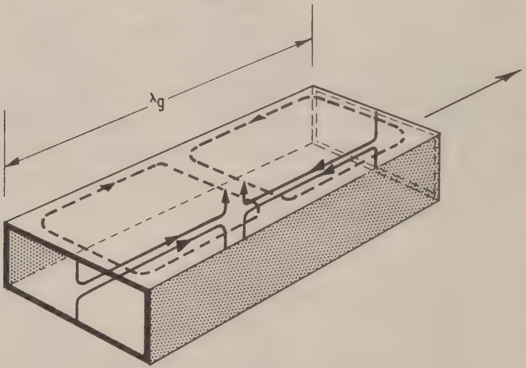


(A) Top view of waveguide. (B) Side view of waveguide.

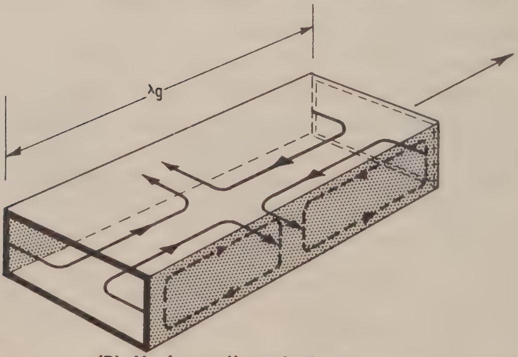
Fig. 3-26. Paths of two-wave patterns.

26B shows them from the side. Since each pattern is composed by two sets of electromagnetic waves which are reflected by the top and bottom of the waveguide, the pattern as a whole is reflected and it behaves, generally, like one of the sets of waves in a $TE_{m,o}$ mode.

Each of the patterns formed by combining two sets of waves has a $\lambda_a/\cos B$, where the wavelength is measured in a direction parallel to the path of the pattern. The pattern moves along the path (Fig. 3-26) with a velocity $\lambda_c/\cos B$, energy moves along the path with a velocity $c \cos B$. The wavelength and the two velocities are equal to the guide wavelength and the phase and group velocities of the $TE_{o,n}$ mode. As might be expected, the patterns cannot exist unless λ_a is shorter than $2b/n$, the cut-off wavelength for the mode. The two patterns combine in the same way as the two sets of waves that form $TE_{m,o}$ mode, and angle A must satisfy the equation



(A) Vertically polarized waves.



(B) Horizontally polarized waves.

Fig. 3-27. Configuration of the combined pattern.

$$\sin A = m\,\lambda/2a,$$

where λ is the wavelength of each of the patterns, or $\lambda_s/\cos B$.

Field Configuration of Complete Pattern

The general appearance of the combined pattern is shown in Fig. 3-27, where the direction of the electric field is shown by solid lines and that of the magnetic field shown by dashed lines. The pattern for vertically polarized waves is indicated in Fig. 3-27A, and the pattern for horizontally polarized waves is shown in Fig. 3-27B. In each pattern, both *m* and *n* are equal to 1. The pattern for higher values of *m* and *n* is simply a repetition of the simple pattern, either up or down or across the guide, as in Figs. 3-24 and 3-25.

To simplify the diagrams, most of the lines that show field direction have been omitted. The direction of the electric field is shown in a plane through the center of the guide, a vertical plane for the vertically polarized waves, and a horizontal plane for the horizontally polarized waves. Similarly, the direction of the magnetic field is shown only near the top of the waveguide and near the right-hand wall. The pattern of the magnetic field near the bottom of the waveguide is the same as the pattern near the top, with the directions reversed. A similar situation with identical patterns and reversed directions exists for the magnetic fields near the two side walls.

$TE_{m,n}$ and $TM_{m,n}$ Modes

If eight sets of electromagnetic waves are present in a rectangular waveguide, both of the patterns shown in Fig. 3-27 will be present. If the phase relation between the two patterns is the same as that in the drawing, the longitudinal component of the magnetic field will be cancelled and the result will be a $TM_{m,n}$ mode. If, on the other hand, the phase of one of

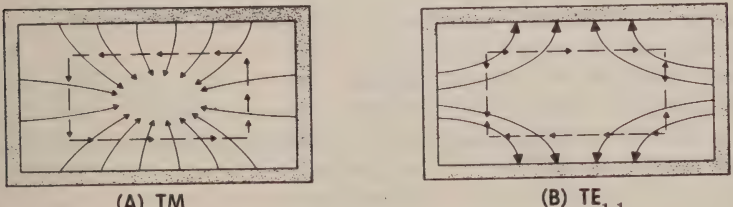

(A) $TM_{1,1}$ (B) $TE_{1,1}$

Fig. 3-28. $TM_{1,1}$ and $TE_{1,1}$ modes in rectangular waveguide.

the patterns is reversed (by reversing the phases of the four sets of waves that contribute to it), the longitudinal electric field will be cancelled, and a $TE_{m,n}$ mode will result. Figs. 3-28A and B show the direction of the resulting electric and magnetic fields in a plane across the waveguide at a point where the transverse field is at a maximum.

Single-Mode Operation of Rectangular Waveguide

Assume that the dimensions of a rectangular waveguide are chosen so that λ_c is shorter than λ_a for the $TE_{1,0}$ mode. The waveguide then cannot carry energy at λ_a by means of the $TE_{1,0}$ mode and this mode is said to be suppressed for λ_a and the corresponding frequency. The cut-off wavelength becomes shorter as either m or n is made larger. Consequently, *all* modes for which m is 1 or a larger number will be suppressed. (It is not necessary to specify the values of n because *any* value is equal to zero or a larger number.) The only possible modes are, therefore, those for which m is zero. Since $TM_{0,n}$ modes do not exist, these are the $TE_{0,n}$ modes. If, in addition, the $TE_{0,2}$ mode is suppressed, all modes for which n is 2 or greater will also be suppressed. There remains only the $TE_{0,1}$ mode (since neither the $TE_{0,9}$ nor the $TM_{0,0}$ mode exists). This procedure normally is followed with rectangular waveguide. Height a is made less than $\lambda_a/2$ so that the $TE_{1,0}$ mode, which has a cut-off wavelength of 2a, is suppressed. The $TE_{0,2}$ mode, which has a cut-off wavelength of b, is suppressed by making b less than λ_a. The cut-off wavelength for the $TE_{0,1}$ mode is 2b, and if the width is more than $\lambda_n/2$ (but less than λ_a), the waveguide will operate in this mode only.

RECTANGULAR-WAVEGUIDE MODES

In microwave operation, it is desirable to use a waveguide that is as large as possible, provided that undesired modes are suppressed. A large waveguide has less attenuation than a small waveguide, just as a large wire has smaller losses (because of its lower resistance) than a small wire when both carry the same current. Modes other than the $TE_{0,1}$ mode are suppressed for much the same reason that standing waves on a transmission line are suppressed. Both absorb power, and both interfere with the proper operation of the system by changing impedance. It is usual to use 1.5- by 3-inch rectangular waveguide for S-band operation and ½- by 1-inch rectangular waveguide for X-band operation. Some of the characteristics of these two rectangular-waveguide sizes are tabulated in Table 3-1.

In practice, the shortest operating wavelength is 5 or 10% longer than the wavelength at which double mode operation is possible. This insures that the waveguide will operate only in the $TE_{0,1}$ mode even when there is a small error in the operating frequency. The longest operating wavelength is about 80% of λ_c for the $TE_{0,1}$ mode. The waveguide will operate at longer wavelengths (so long as they are shorter than λ_c), but the attenuation increases rapidly as the operating wavelength approaches the cut-off wavelength. A good rule of thumb for estimating the operating wavelength of an odd size of rectangular waveguide is that the nominal operating wavelength is about five-fourths of the larger dimension.

Table 3-1. Waveguide Characteristics

	Nominal operating wavelength	
	(S-band)[a]	(X-band)[b]
Outside dimensions in inches	1.500 x 3.000	.500 x 1.000
Wall thickness in inches	.080	.050
Inside dimensions in inches	1.340 x 2.840	.450 x .950
Inside dimensions in centimeters	3.40 x 7.22	1.14 x 2.41
$\lambda_c TE_{0,1}$ mode	14.44 cm	4.82 cm
$\lambda_c TE_{0,2}$ mode	7.22 cm	2.41 cm
$\lambda_c TE_{1,0}$ mode	8.80 cm	2.28 cm

[a] S-band—10 cm
[b] X-band—3 cm

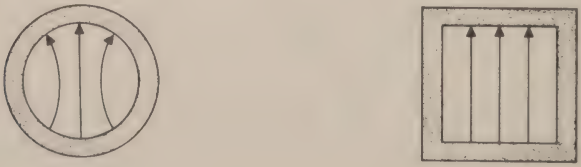
CIRCULAR WAVEGUIDE MODES

The behavior of circular waveguide is similar to that of rectangular waveguide. The field patterns, like those of rectangular waveguide, satisfy the following two boundary conditions imposed by the fact that the waveguide wall is a good conductor: At the surface of a wall there can be no electric field parallel to the wall, and there can be no varying magnetic field perpendicular to the wall. A review of the material on rectangular waveguides will show that these boundary conditions are satisfied by the field patterns of all existing modes. Because of the curved walls of circular waveguide, it is extremely difficult to examine the field patterns in the same detail as those of rectangular waveguide—there is no simple method of building the field pattern from a small number of simple waves traveling along the guide. The calculation of cut-off wavelength for a given mode in circular waveguide is, therefore, beyond the scope of this book, and it will be necessary to accept the values given here without any demonstration of their validity. The general configuration of the field patterns may be deduced, however, from their similarity to corresponding patterns in square waveguide (rectangular waveguide for which *a* and *b* are equal).

As with rectangular waveguide, all circular waveguide modes are either TE, TM, or combinations of these (TE and TM have the same significance as they have in rectangular waveguide modes). The modes are described further by the same two subscripts so that there are $TE_{m,n}$ modes an $TM_{m,n}$ modes. In rectangular waveguide, *m* is the number of times the field pattern is repeated from the top of the guide to the bottom, with a reversal in the directions of the electric and magnetic fields at each repetition; and *n* is the number of times the pattern is repeated from one side wall to the other. In circular waveguide, however, both *m* and *n* have different meanings: *m* is the number of times the pattern is repeated around the waveguide, and *n* is the number of times the pattern is repeated between the center of the guide and the wall.

Important Modes

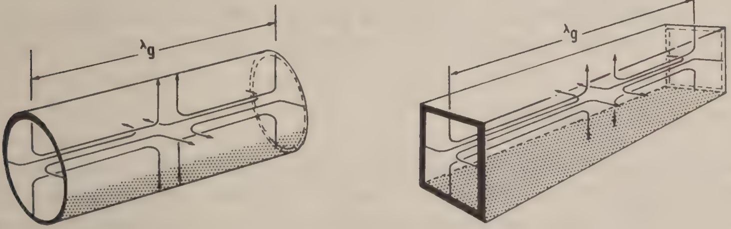
In considering the behavior of circular waveguide, it is necessary to examine only the $TE_{1,1}$ $TM_{0,1}$ and $TE_{0,1}$ modes. These are the three simplest modes for a circular waveguide



(A) $TE_{1,1}$ (B) $TE_{0,1}$

Fig. 3-29. Dominant TE modes in circular and square waveguide.

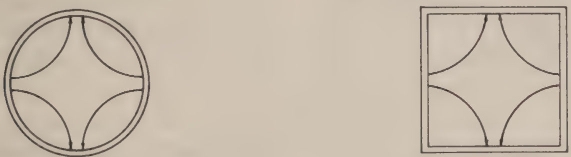
and the only ones for which the cutoff wavelength is greater than the diameter of the guide. $TE_{1,1}$ is the dominant mode (that having the longest cutoff wavelength) for a circular waveguide; it has a cutoff wavelength of 1.71d, where *d* is the inside diameter of the waveguide. Fig. 3-29A shows the electric field pattern for the $TE_{1,1}$ mode, and Fig. 3-29B shows the electric field for the $TE_{0,1}$ mode in a square waveguide (a rectangular guide for which *a* and *b* are equal). The resemblance between the fields in the circular and square waveguides is obvious, and both modes are dominant modes. The $TE_{2,1}$ mode with a cutoff wavelength of 1.03d is shown in Fig. 3-30A; and the $TM_{1,1}$ mode in a square waveguide is shown in Fig. 3-30B. The two diagrams are simplified by omitting the lines showing the direction of the magnetic field. The $TE_{2,1}$ mode with a cutoff wavelength of 1.03d is shown in Fig. 3-31A. The corresponding mode for a square waveguide is the $TE_{1,1}$ mode (Fig. 3-31B). All other modes, both TE and TM, have cutoff wavelengths of 0.82d or less.



(A) $TM_{0,1}$ (B) $TM_{1,1}$

Fig. 3-30. Simplest TM modes in circular and square waveguide.

The usefulness of circular waveguide lies in the perfect symmetry of the $TM_{0,1}$ mode. A rotating joint can be inserted in a section of circular waveguide without affecting the transmission of energy. Therefore, a section of circular waveguide is used to join two sections of rectangular waveguide when



(A) $TE_{2,1}$ (B) $TE_{1,1}$

Fig. 3-31. Simplest TE modes in circular and square waveguide.

it is necessary to move one in relation to the other (Fig. 3-32). Moving the upper rectangular waveguide from the position shown in *A* to the position shown in *B* has no effect on the transmission of energy. Although it is possible to use the $TE_{1,1}$ mode of circular waveguide, it is difficult to control the signal-polarization plane. Thus, the corresponding $TE_{0,1}$ mode in rectangular waveguide is normally used instead.

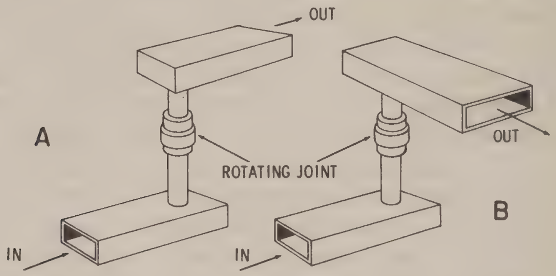


Fig. 3-32. Circular waveguide rotating joint.

Waveguide Dimensions

Circular waveguide is always chosen so that λ_a is shorter than 1.31d, the cutoff wavelength for the $TM_{0,1}$ mode. The higher TM modes then are automatically suppressed by choosing the diameter so that λ_a is longer than 0.82d. The method of exciting the waveguide suppresses the $TE_{1,1}$ and $TE_{2,1}$ modes unless the second of these already is suppressed by a choice of diameter. It is usual to excite the waveguide with a small antenna at the guide axis (Fig. 3-33). The field radiated by this antenna has no longitudinal magnetic component (which must be present for any TE mode), and, consequently, neither the $TE_{1,1}$ nor the $TE_{2,1}$ mode is excited. Mechanically, it is desirable to use circular waveguide with a diameter equal to the width of the rectangular waveguide in the system so that *d* is normally about $0.8\lambda_a$, but this is not always done.

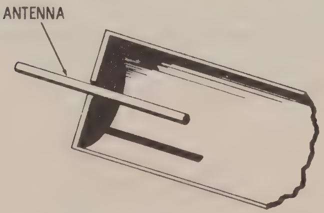


Fig. 3-33. Method used to excite the waveguide.

The circular waveguide normally used in S-band systems has a diameter of about 3 inches, and that used in X-band systems is about 1 inch in diameter.

WAVEGUIDE ADVANTAGES

The major advantage of waveguide over coaxial line is the much lower loss in waveguide, resulting primarily from the elimination of the center conductor. At VHF and microwave frequencies, interaction between the guided waves in a transmission system and the current-carrying electrons causes the current flow to be confined to the surface of the conductors. This *skin effect,* of course, is present at lower frequencies, but is unimportant. However, at VHF and microwave frequencies, because of the skin effect, only the surface of the inner conductor of coaxial line is available for current carrying, and its resistance is unavoidably high. The outer conductor, having a larger surface (the inner surface is the one that carries the current) has a much lower resistance. The major portion of the loss, therefore, occurs at the inner conductor of coaxial line, which is not present in waveguide and is not present when waveguide is used. Waveguide has other advantages, such as lack of radiation and dielectric losses, but rigid coaxial line offers the same advantages. The greatest merit of waveguide, aside from the fact that it has lower conductor loss (sometimes called copper loss) than coaxial line, is its extreme simplicity of construction. By comparison with waveguide, a rigid coaxial line with quarter-wave stub supports is a complicated mechanical device.

TUNING DEVICES AND RELATED EQUIPMENT

Before considering the accessories used with waveguide, it is desirable to examine some of those used with coaxial line at microwave frequencies because their operating principles are more familiar than those of the waveguide components. The coaxial-line accessories are specialized to a certain extent because of the line construction itself. In general, they are rigid structures which, like the coaxial line, avoid the use of any insulating material other than dry air. Several of the more usual components are tuning stubs, slotted sections, line stretchers, and rotating joints.

Tuning Stubs

A tuning stub is a short section of coaxial line connected across the main line. The arrangement is similar to that of the quarter-wave stub support, except that the length of the tuning stub can be adjusted to any value between zero and one-half wavelength. Two tuning stubs are usually mounted together on a single section of coaxial line inserted in the main transmission line, the spacing between the two stubs usually being three-eights wavelength (Fig. 3-34).

The inner and outer conductors have the same dimensions as those of the remainder of the coaxial line, and a single quarter-wave stub support holds the inner conductor of the section in its proper place. A flange, provided at each end of the section, makes contact with the outer conductor of the adjacent portion of the line. The inner conductor has a socket at one end and a finger at the other end; these

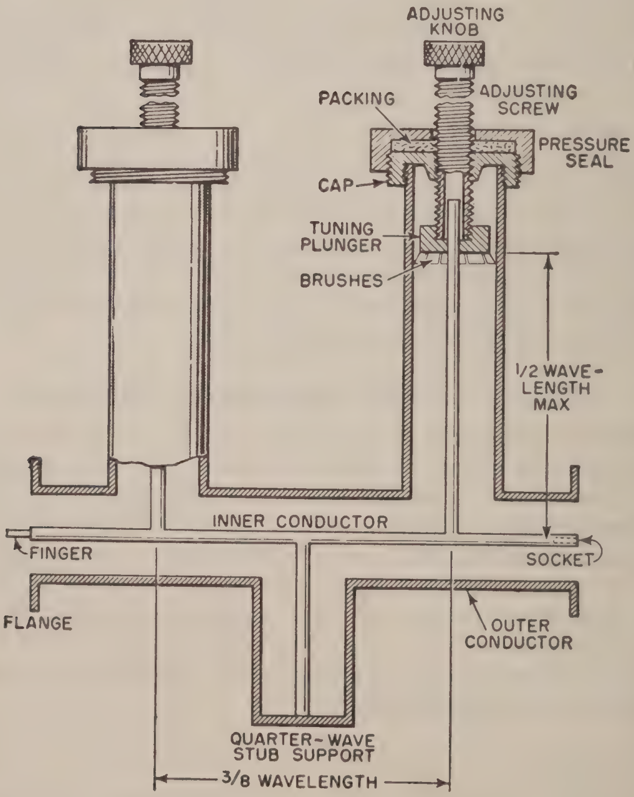

ADJUSTING KNOB
PACKING
ADJUSTING SCREW
PRESSURE SEAL
CAP
TUNING PLUNGER
BRUSHES
1/2 WAVE-LENGTH MAX
INNER CONDUCTOR
FINGER
SOCKET
FLANGE
OUTER CONDUCTOR
QUARTER-WAVE STUB SUPPORT
3/8 WAVELENGTH

Fig. 3-34. Double-stub section of coaxial line.

parts connect with matching parts on adjacent sections of the line and provide continuity of the inner conductor. The finger usually is split and slightly spread apart to make a good friction fit in the matching socket. The two stubs are identical. Each consists of an inner and outer conductor and an adjustable plunger which acts as a short circuit. The length of the stub is varied by changing the plunger position.

A number of small metal brushes are provided on the plunger to insure good contact with both the inner and outer conductors. The plunger is carried on one end of a hollow screw which has an adjusting knob at the other end; the screw itself is carried in a threaded cap mounted on the end of the outer conductor. A pressure seal with packing under it serves to prevent loss of air from the coaxial line and also acts as a lock for the adjusting screw. There are, of course, many different ways of constructing a double-stub section, but the one shown in Fig. 3-34 is a good illustration of the principles involved in the types in general use.

Theory

Every transmission line has a characteristic impedance, usually written as Z_0, which depends on the physical characteristics of the line. If a current (I) flows in the line, the voltage (E) accompanying it then is derived from the equation:

$$E = Z_0 I$$

This equation holds for signals traveling in either direction along the line so that

$$E_1 = Z_0 I_1$$

And, by division,

$$I_1 = \frac{E_1}{Z_0}$$

and

$$E_2 = Z_0 I_2, \text{ or } I_2 = \frac{E_2}{Z_0}$$

where, E_1 and I_1 are the voltage and current carrying power from a generator to a load, and E_2 and I_2 are the voltage and current of the reflected signal carrying power in the opposite direction. Fig. 3-35 shows the positive direction for E_1, E_2, I_1, and I_2. For convenience, the transmission line is shown as a two-wire line since the discussion applies to this type as well as to coaxial line. It can be seen that the voltage measured at any point will be the sum $(E_1 + E_2)$ and the current will be the difference $(I_1 - I_2)$ when the phase relations between E_1 and E_2, and between I_1 and I_2 are taken into account.

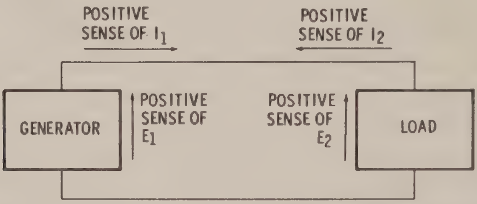

POSITIVE SENSE OF I_1
POSITIVE SENSE OF I_2
GENERATOR
POSITIVE SENSE OF E_1
POSITIVE SENSE OF E_2
LOAD

Fig. 3-35. Directions of direct and reflected signals in transmission line.

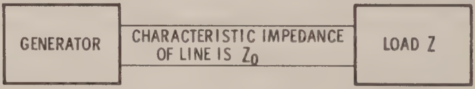

GENERATOR
CHARACTERISTIC IMPEDANCE OF LINE IS Z_0
LOAD Z

Fig. 3-36. Generation of the reflected signal.

If the transmission line is terminated by a load of impedance *Z* (Fig. 3-36) and if *E* and *I* are the voltage and the current at the load, E/I must be equal to Z. In this case,

$$Z = \frac{E}{I} = \frac{(E_1 + E_2)}{(I_1 - I_2)}$$

But, since $I_1 = E_1/Z_0$, and $I_2 = E_2/Z_0$, these values can be substituted in $(I_1 - I_2)$,
and, the equation

$$E_2 = \frac{E_1 (Z - Z_0)}{(Z + Z_0)}$$

is obtained by simple algebra. If Z is equal to Z_0, $(Z - Z_0)$ will be equal to zero, and E_2 (and consequently I_2) also will be equal to zero. In simpler terms, there is no reflected signal when a transmission line is terminated in its characteristic impedance. The reflected signal is objectional for two reasons, the first of which has to do with power losses on the transmission line. Suppose that the loss on a given line is 10%. An input of 100 watts then will provide 90 watts at the load if there is no reflected signal. If, however, there is a reflected signal, more input power is required to deliver the same 90 watts to the load. Suppose, for example, that E_2 is one-half of E_1. I_2 also will be one-half of I_1, and the power in the reflected signal will be one-fourth of the power in the direct signal.

If 120 watts arrives at the load end of the transmission line, 30 watts will be reflected and 90 watts will be delivered to the load. In order to deliver 120 watts at the load end, however, it is necessary to put 133 1/3 watts in at the generator end of the line because of the 10% transmission line loss. This extra power requirement is partially offset by the reflected signal, which delivers 27 watts (because of the 10% transmission line loss) at the generator end. The generator itself is then called on to supply 106 1/3 watts (the difference between 133 1/3 and 27 watts) in order to develop 90 watts at the load. This is an increase of 6 1/3 watts because of the reflected signal, and any power increase is undesirable.

The second objection to the reflected signal is the possibility of an impedance mismatch at the generator. If Z is equal to Z_o, there is no reflected signal, and the voltage and current at the input end of the transmission line will have the ratio equal to Z_o. The generator customarily is designed to deliver maximum power output when it sees this impedance. In the presence of a reflected signal, however, the impedance seen by the generator is equal to the ratio $(E_1 + E_2)/(I_1 - I_2)$. This ratio is, of course, equal to Z at the load end. It is necessary, however, to take account of phase when adding and subtracting the voltages and currents. For a position between the generator and the load, the phases of E_1 and I_1 are advanced, and those of E_2 and I_2 are retarded. Because of this, the impedances measured at different points on the line have different values, and, in particular, the impedance measured at the generator may have values as far apart as Z an Z_0^2/Z. With proper matching, the generator in the example above may develop 200 watts, of which 100 watts is delivered to the transmission line, and the rest is lost in the generator. When there is a reflected signal of the strength given above, the generator is required to supply 105 1/3 watts to the transmission line. An impedance mismatch resulting from a change in the input impedance of the line, however, may require that the generator develop 400 watts in order to transfer 106 1/3

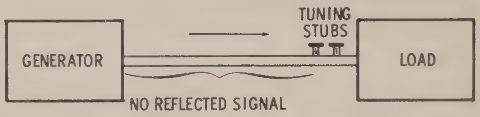


Fig. 3-37. Tuning stubs used to cancel reflected signals.

watts to the line. The tuning stubs shown in Fig. 3-37 are used when the load impedance is not equal to the characteristic impedance of the line. They are placed on the line close to the load and are adjusted so that the reflected signals from the stubs cancel the reflected signal from the load. The portion of line between the stubs and the generator is therefore free of reflected signal. Each stub delivers a reflected signal because it is a section of line terminated in a zero impedance.

Adjustment—In practice, adjustment of the tuning stubs is started by setting both stubs at their midpositions. Each is then a quarter-wavelength and looks like an infinite impedance across the line (so that the stubs have no effect). The stub nearest the generator then is moved, while the load power is observed. An initial adjustment is reached for which the load power is a maximum. (In a radar, this may be accomplished by observing a fixed echo and adjusting the stub for maximum echo strength.) After the initial adjustment, the second stub, nearer the load, is moved a little way and the first stub is readjusted for maximum power. Depending on whether this is an improvement or not, the second stub is then moved farther in the same direction, or back in the opposite direction. The process continues until the best adjustment has been found by trial, always with a readjustment of the first stub after the second one has been moved.

Slotted Section

When it is not convenient to observe the load power while the tuning stubs are being adjusted, it is necessary to make use of a slotted coaxial-line section (Fig. 3-38) so that the voltage on the transmission line may be measured at various points. The measured voltage is the sum of E_1 and E_2, where these are added vectorially because of the phase difference between them. At some point, the two voltages will be in phase and the measured voltage will have a maximum value.

At another point, the two voltages will be 180° out of phase and the measured voltage will have a minimum value. Then, if $E_{1\text{-peak}}$ is the peak value of E_1 and $E_{2\text{-peak}}$ is the peak value of E_2, with E_{max} and E_{min} the maximum and minimum values of the measured voltage:

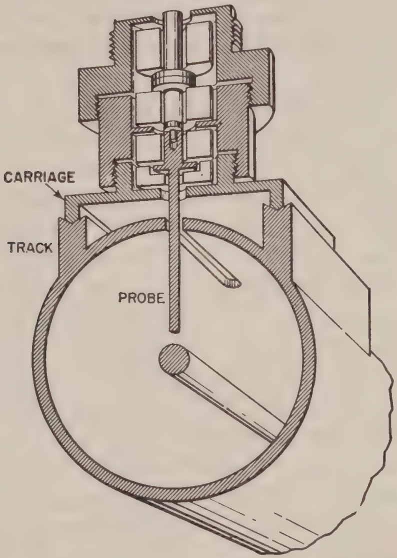


Fig. 3-38. Slotted section of coaxial line.

$$E_{max} = E_{1\ peak} + E_{2\ peak}$$

$$E_{min} = E_{1\ peak} - E_{2\ peak}$$

The ratio E_{max}/E_{min} is called the VSWR (voltage standing wave ratio) and it is a measure of the amount of reflected signal present. If the VSWR is one (1), there is no reflected wave. The adjustment of the tuning stubs may then be accomplished by observing the VSWR instead of the power in the load and adjusting for the lowest VSWR. Values of three usually are considered acceptable in operation at microwave frequencies, although lower values are desirable if they can be obtained.

The slotted section is simply a section of line with a slot in the outer conductor so that the inner conductor becomes accessible. A probe is mounted on a small carriage, which moves on a track fixed to the outer conductor. The probe does not touch the inner conductor, and a signal is picked up by virtue of the capacity between the inner conductor and the tip of the probe. It is important that the spacing between the probe and the center conductor remain constant as the probe is moved, and the carriage and track must be inspected from time to time. The signal picked up by the probe is rectified and sent to an indicator of some sort, usually a milliammeter which is preceded by an amplifier. The VSWR is found by moving the carriage back and forth and noting the maximum and minimum indicator readings, the ratio of these being the VSWR.

The slotted section may be quite short. If the two voltages add to produce a maximum at one point, they will do so at other points that are spaced an integral number of half-wavelengths away. This occurs because a position change of one-half wavelength shifts the phase of E_1 by one-half cycle in one directon and the phase of E_2 by one-half cycle in the other direction, thus leaving them in the same phase relation as before. Midway between each two adjacent points where the measured voltage is a maximum, there is a point where E_1 and E_2 are 180° out of phase, and the measured voltage is minimum. It is always possible, then, to find one maximum and one minimum if the slot is one-half wavelength long. In practice the slot is longer than this, and several readings are taken at different maxima and minima.

A slotted section frequency is built in as part of a transmission line. If, however, the line is to be pressurized, it is necessary to provide a removable section of ordinary coaxial line which can be replaced by the slotted section when the VSWR is to be measured.

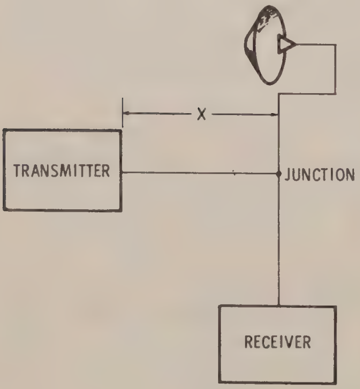


Fig. 3-39. An installation with precision transmission line length.

Line Stretcher

In some installations a line is terminated by an impedance different from Z_o and the length of the line is adjusted so that the input impedance has a specified value (Fig. 3-39). In this example, if length X is adjusted properly, the impedance looking toward the transmitter from the junction will be infinite when the transmitter is turned off (between pulses). The transmitter, therefore, does not take any energy from the received signal. Because of frequency variations in different transmitters, it is not practical to adjust length X at the factory. Instead, a device called a line stretcher (Fig. 3-40) is installed between the transmitter and the junction as part of the transmission system. Its operation is obvious, since it consists of a U-shaped section which slides up and down to change the length of the line. It is the coaxial equivalent of a slide trombone.

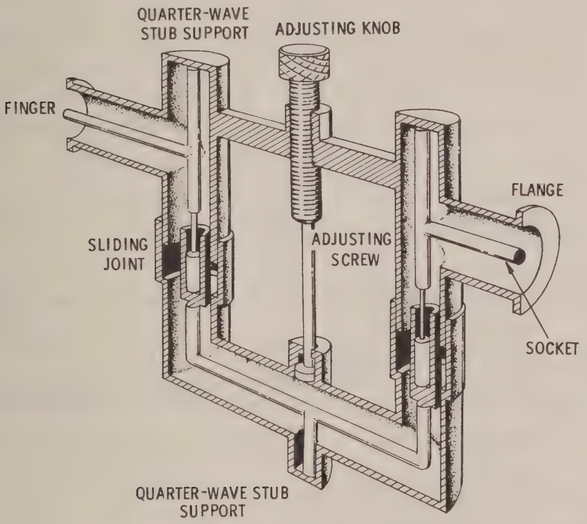


Fig. 3-40. Coaxial-line stretcher.

Rotating Joint

It is frequently necessary for one part of a transmission line to move while another part remains fixed. When the system consists of rigid coaxial line, this freedom is obtained by using a rotating joint. A of Fig. 3-41 shows a quarter-wave

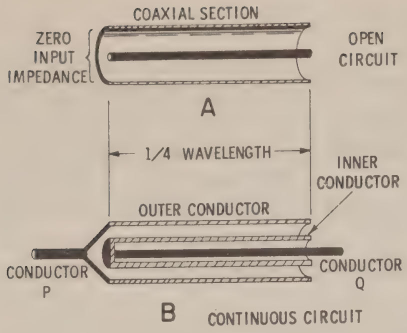

COAXIAL SECTION
ZERO INPUT IMPEDANCE
OPEN CIRCUIT
A
1/4 WAVELENGTH
INNER CONDUCTOR
OUTER CONDUCTOR
CONDUCTOR P
CONDUCTOR Q
B
CONTINUOUS CIRCUIT

Fig. 3-41. Quarter-wave coaxial section used for electrical connection with mechanical freedom.

section of coaxial line terminated by an open circuit. This section has zero input impedance; in other words, it acts like a short circuit at its input end. In practice, some capacity always exists at the terminating end and the section must be made slightly less than a quarter wavelength. In B of Fig. 3-41, the same section is shown with the inner conductor made hollow. It is then possible to bring a pair of conductors, *P* and *Q*, to the sending end of the quarter-wave section and provide a good electrical connection between them without any mechanical contact. In Fig. 3-42, B shows this device used to make a rotating joint in coaxial line, and A shows a practical joint developed from the scheme used in B. Such a joint is used frequently but is difficult to seal because of the high voltage between M and N. It is therefore modified to the form shown in Fig. 3-43 when the system is pressurized. It is possible to make a seal between *U* and *V* without difficulty, since

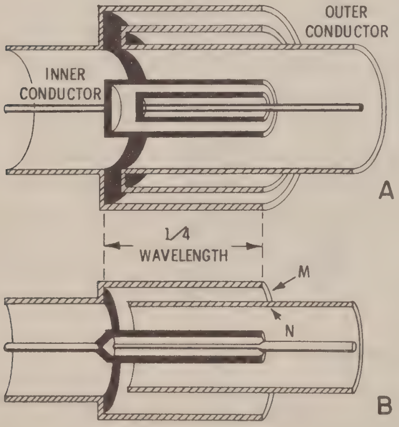

OUTER CONDUCTOR
INNER CONDUCTOR
A
1/4 WAVELENGTH
M
N
B

Fig. 3-42. Rotating joints.

these points can be connected without affecting the electrical operation. The resistance of the seal is in series with the infinite input impedance of the shorted quarter-wave section to the right of V, and, whether the seal is good or bad from an electrical standpoint, the quarter-wave section to the left of U is always terminated in the required open circuit. Through the use of a half-wave section the connection for the outer conductor is terminated in a short circuit, which also has zero input impedance.

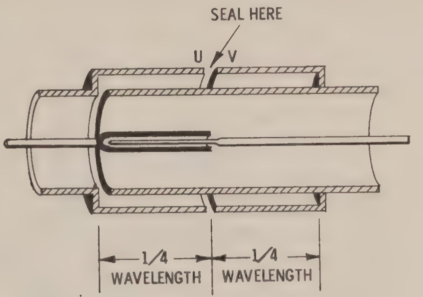
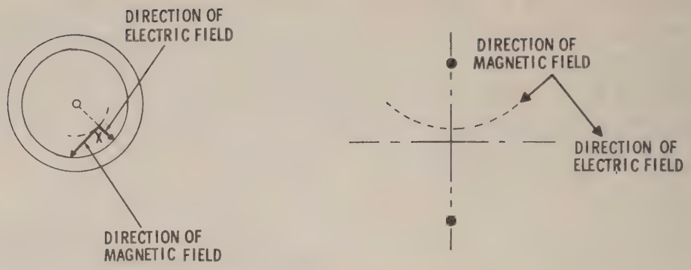


Fig. 3-43. Pressurized rotating joints.

Waveguide Tuning Devices

At first glance it seems inappropriate to speak of the impedance of a waveguide, since there is no point at which the voltage and the current can be measured. The concept of waveguide impedance is, however, as useful as that of transmission line impedance.

Waveguide Impedance—The impedance of a transmission line at any point is the ratio between the voltage and the current at that point. It is necessary, of course, when computing this ratio, to take into account the phase relationship between the voltage and current. The impedance of a waveguide at any point is, similarly, the ratio between the transverse magnetic field at that point. As with transmission line impedance, it is necessary, when computing the ratio, to take into account the phase relationship between the two fields. Fig. 3-44A shows the fields in a coaxial line, and Fig. 3-44B shows the fields in a conventional two-wire line. In both the transmission line acts as a guide for the guided waves that accompany any transmission of energy. At any point where the guided wave is present, such as X in Fig. 3-44A and Y in Fig. 3-44B, electric and magnetic fields are associated with the guided waves.


DIRECTION OF ELECTRIC FIELD
DIRECTION OF MAGNETIC FIELD
DIRECTION OF MAGNETIC FIELD
DIRECTION OF ELECTRIC FIELD

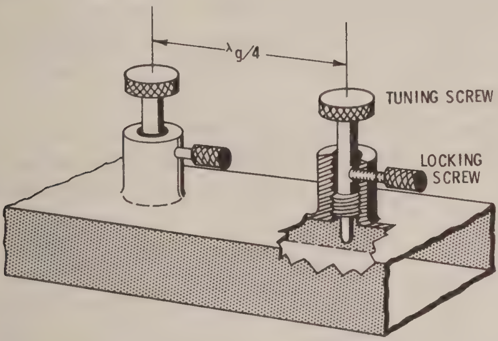
(A) Coaxial. (B) Two-conductor.

Fig. 3-44. Associated fields of transmission lines.

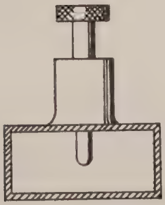
The transverse component of the electric field (the field in a plane perpendicular to the conductors) depends on the position of the point and on the voltage between the conductors. For any given point the strength of the electric field is proportional to the voltage. In the same way, the transverse magnetic field depends on the position of the point and on the current carried by the conductors. The strength of the magnetic field is proportional to the current for any given point. Both the transverse electric and magnetic fields depend in the same way on the position of the point, so that the ratio between them does not depend on the position of the point. The ratio between the two transverse fields is proportional to the ratio between the voltage and the current; it follows that the impedance of a transmission line can be defined either by the ratio between the voltage and the current, or by the ratio between the transverse electric and transverse magnetic fields.

The second way of defining the impedance can be applied directly to a hollow waveguide. The waveguide, then, has a characteristic impedance associated with each mode (although several modes may have the same characteristic impedance). As explained previously, the actual impedance is equal to the characteristic impedance when no reflected wave is present. In the presence of a reflected wave the impedance is different from the characteristic impedance and is different at different points along the waveguide. The impedance is always the same at all points in a plane perpendicular to the axis of the waveguide, just as the impedance of a transmission line (computed from the electric and magnetic fields instead of from voltage and current) is the same no matter what point in a given plane is chosen to measure the fields.

Slug Tuner—A slug tuner is used with a waveguide transmission system in precisely the same manner that the stub tuner is used with a conventional transmission line (the slug tuner is essentially the same device as the stub tuner). Most waveguide systems consist largely of a rectangular waveguide operating in the $TE_{0,1}$ mode; the double-slug tuner shown in Fig. 3-45 is designed for this type of waveguide. Fig. 3-45A



(A) Side view.



(B) End view.

Fig. 3-45. Double-slug tuner for the $TE_{0,1}$ mode in rectangular waveguide.

shows a section through the tuner, and Fig. 3-45B shows an end view. Each tuning slug projects into the waveguide parallel to the electric field and acts as a small antenna. It is excited by the electric field in the waveguide and consequently radiates a signal. The phase of the reflected signal may be altered by changing the length of the slug with the tuning screw, and the signals reflected from the slugs cancel the signals reflected from the load.

Electrically, the double-slug tuner is similar to the double-slug tuner shown in Fig. 3-34, with one exception: the range over which the stubs can be varied is less than the range over which the slugs can be varied. The tuning stub has no effect when it is a quarter wavelength, and its length may be varied in either direction from this point. The slug is ineffective when it is backed out to a position even with the waveguide wall, and it can be varied in only one direction from this point. Consequently, the double-slug tuner is limited to some extent, and three-slug tuners frequently are used with spacing of

$3/8\lambda_g$ between the adjacent slugs. A double-slug tuner is adjusted by starting with both slugs retracted so that their tips are even with the wall of the waveguide, the position in which they have no effect. One of the slugs then is inserted by advancing the tuning screw. If this makes the situation worse, that slug is retracted and the other is advanced. From this point, tuning proceeds by alternate adjustment of the two slugs. This method is similar to the one described for the double-stub tuner.

Tuning Stubs—Although actual tuning stubs are occasionally used, in most applications they offer little advantage over the slug tuner. A typical tuning stub is shown in Fig. 3-46.

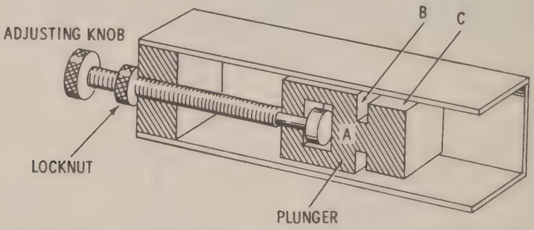
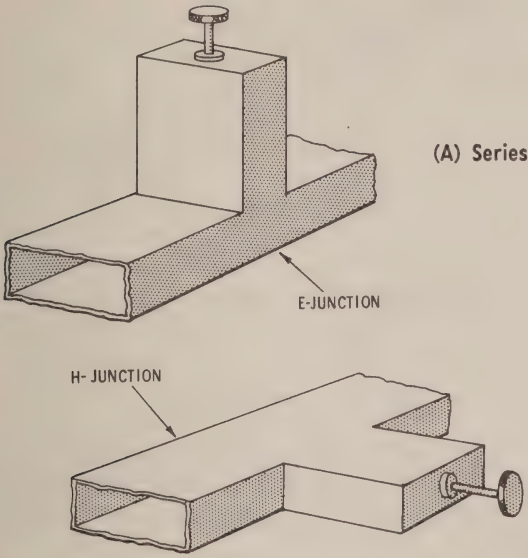


Fig. 3-46. Adjustable tuning stub.

It consists of a waveguide section with an adjustable plunger of conducting material. To insure good contact between the plunger and the wall of the waveguide and, therefore, an effective short circuit for the termination, the plunger is grooved a short distance back from its face. The groove, when seen from *B* in the figure, looks like a short section of transmission line terminated by the short circuit at *A*. The groove depth is chosen so that this section of line has an effective length of one-quarter wavelength. The impedance at *B* is therefore the infinite impedance of a shorted quarterwave section of line. This is in series with the impedance of the rubbing contact between the plunger and the waveguide wall to the left of *B*. Distance BC is also one-quarter wavelength, and the impedance of the gap between the plunger and the wall at *C* is that of a quarter-wave section terminated by the impedance at *B*. Since the impedance at *B* is always great (because of the impedance of the groove), the impedance at *C* is almost zero, whether or not there is a good contact between the plunger and the wall. The groove in the plunger is called a *choke groove*. The stub may be added to either wall of the waveguide. When the junction is at the wall perpendicular to the electric field, it is called an *E-junction*, the stub is called a *series stub*, and the impedance of the stub is in series with that of the main line. When the junction is at the wall perpendicular to the magnetic field, it is called an *H-junction*, the stub is called *shunt stub*, and the impedance of the stub is in parallel with that of the main line. Fig. 3-47A shows a stub



(A) Series stub.

(B) Shunt stub.

Fig. 3-47. Shunt and series tuning stubs.

connected with an E-junction, and, in Fig. 3-47B, the stub is connected with an H-junction. (In practice, the E-junction is the one normally used.)

Window Tuners—It is always possible to tune a waveguide by any method that introduces the required reflection, and this sometimes is done by means of window tuners (Fig. 3-48). Since it is desirable for the two obstructions to be symmetic, it is rather difficult to construct an adjustable window tuner; however, these tuners are useful whenever they can be preset and left alone. They are electrically equivalent to shunt tuning stubs. The window tuners are nothing more than small conducting plates inserted in the guide. The tuner shown in A is located where the electric field is relatively weak, and its major effect is on the magnetic field. Since the magnetic field is analogous to current in the impedance of a waveguide, a tuner of this sort behaves like a current device and is approximately the equivalent of a shunt inductor (which does not

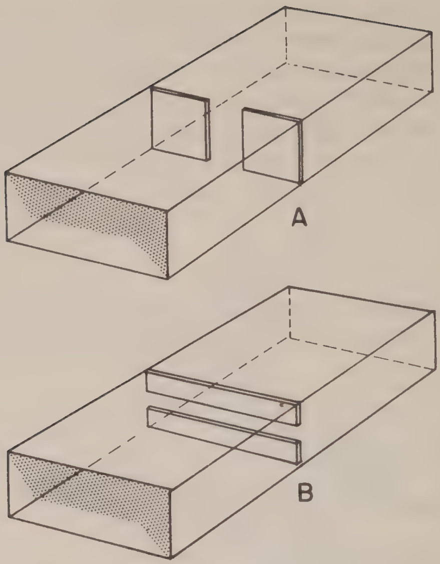


Fig. 3-48. Window tuners.

affect the voltage) on a conventional transmission line. The tuner shown in B affects the electric field more strongly than the magnetic field and is the approximate equivalent of a shunt capacitor.

Slotted Sections—Whenever a waveguide carries both direct and reflected signals, the strength of the electric field varies

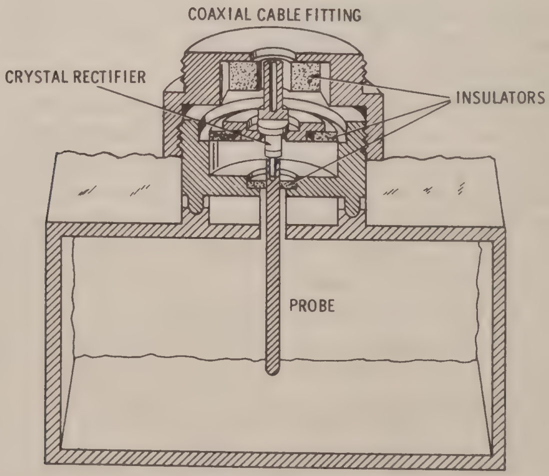

COAXIAL CABLE FITTING
CRYSTAL RECTIFIER
INSULATORS
PROBE

Fig. 3-49. A slotted section.

from point to point along the guide, just as voltage varies in a conventional transmission line. A slotted section with coaxial line is used to observe the load power, as it is used with a waveguide. Following transmission line terminology, the ratio between the maximum and the minimum electric field strengths is called the VSWR. Fig. 3-49 shows a cross section of the carriage which rides on a slotted section of a rectangular waveguide. The probe, which is parallel to the electric field, is excited by the signal and developes a voltage at the operating frequency. A crystal rectifier is mounted in the carriage, and the output of the crystal is fed to a meter. Generally, it is possible to use a sensitive meter directly, but in some cases an amplifier is required. Fig. 3-49 shows the carriage and crystal holder with a fitting for a coaxial cable.

Directional Couplers—A directional coupler is generally used to measure the reflected signal in a waveguide. The directional coupler requires no manipulation and can be sealed to make a permanent part of a pressurized waveguide system.

Fig. 3-50 shows the basis for the theory underlying the use of the directional coupler. Here, a main waveguide has an auxiliary waveguide coupled to it by means of two small holes. Referring to Fig. 3-50A, it is seen that when a signal travels from left to right in the main waveguide, some energy leaks into the auxiliary guide through the holes at WX and YZ, and that the signal in the auxiliary guide at *Z* is composed of two parts. Some energy leaks through the hole at *WX* and

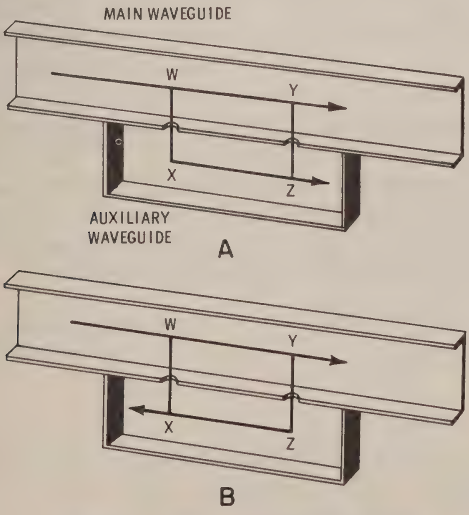


Fig. 3-50. Theory of directional coupler.

travels from *X* to *Z*. Additional energy travels from *W* to *Y* and then leaks through the hole at YZ. Since distance WY and XZ are equal, the two parts of the signal will be in phase and some energy will be propagated from left to right in the auxiliary guide. Propagation of a signal in the opposite direction in the auxiliary guide, as illustrated in B, depends on the spacing between the holes. Some energy leaks directly through the hole at WX. Additional energy travels from *W* to *Y*, leaks through the hole at YZ, and then travels back from *Z* to *X*. The longer path differs from the shorter path by twice the spacing between the holes. If, then, the holes are separated by a distance $\lambda_g/4$, the two signals will be out of phase at *X* and no energy will be propagated from right to left in the auxiliary guide.

Energy from the direct signal leaks into the directional coupler and travels in the same direction to the right-hand end where it is absorbed by a nonreflecting termination which is a resistive surface, the impedance of which matches that of the waveguide. Energy from the reflected signal leaks into the directional coupler and continues to the left-hand end, where it excites the probe. Any reflection from the left-hand termination of the directional coupler reaches the right-hand termination and is absorbed.

The probe is fitted with a crystal holder, as is the carriage of the slotted section shown in Fig. 3-49. The signal picked up by the probe is proportional to the strength of the reflected signal in the main waveguide and does not depend on the level of the signal traveling in the other direction. A waveguide system frequently is fitted with two directional couplers, one to measure the strength of the desired signal and one to measure the level of the reflected signal. It is possible also to feed energy into the directional coupler through the probe. Some of this energy will be absorbed by the resistive termination, and the rest will leak through the two holes into the main waveguide. Because of the difference in the path lengths through the two holes, the energy that reaches the main waveguide will be propagated in only one direction.

Chapter 4

Cavity Resonators

A cavity resonator, sometimes called a resonant cavity, is nothing more than a space bounded by conducting walls—a hollow conductor. Cavity resonators are used at microwave frequencies in place of the LC (inductance-capacitance) circuits and transmission line stubs employed at lower frequencies. The behavior of cavity resonators is discussed in detail in the following paragraphs. Before proceeding to this analysis, however, it is desirable to examine the behavior of a simple cavity in order to gain an understanding of how it resonates.

REINFORCEMENT OF SIGNALS

Consider a section of rectangular waveguide that is excited in the $TE_{0,1}$ mode by a small probe (Fig. 4-1). (The transmission system that supplies energy to the probe is not shown, since it does not enter the discussion.) The waveguide extends a distance, X, to the left of the probe; here, it is terminated

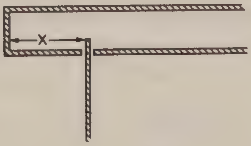


Fig. 4-1. Waveguide terminated by conducting wall.

by a conducting wall and then extends indefinitely to the right of the probe. This probe, of course, radiates signals to the right and to the left. At any instant, the signal moving to the right is made up of two parts: one part is simply the signal that the probe radiates to the right; the other part is the signal that has been reflected from the termination at the left—the signal that was radiated to the left by the probe a short time before. Now, if distance X is equal to $\lambda_g/4$, the signal radiated to the left must travel a half-guide wavelength before it returns to the probe. Its phase is reversed on reflection at the conducting wall so that it reaches the probe with a phase shift of 180°. Because it travels a half guide-wavelength, its traveling time is exactly one-half cycle. During this time the phase of the signal radiated by the probe also changes by 180°. If, therefore, distance X is one-quarter λ_g (the guide wavelength), the signal radiated to the right by the probe will be reinforced by the reflected signal that was originally radiated to the left by the probe traveling to the right. If distance *X* is changed, the traveling time is also changed. The phase shift in the reflected signal, however, is always 180°, since it results from reflection at the terminating wall and not from traveling time. It follows, therefore, that the signal (reflected) will reinforce the signal radiated to the right whenever the traveling time in an odd number of half cycles. The phase shift at the probe during this time is, then, 180° plus some whole number of cycles. Therefore reinforcement occurs whenever 2X is an odd number of half guide-wavelengths, or whenever *X* is an odd multiple of $\lambda_g/4$.

SIMPLE RESONATOR

Now, consider the resonant cavity shown in Fig. 4-2. This consists of a section of waveguide that is similar to the one discussed, except that it is terminated in both directions by conducting walls. If, now, both *X* and *Y* are odd multiples of $\lambda_g/4$, signals radiated to the right will be reinforced by those reflected from the left-hand wall, and signals radiated to the left will be reinforced by those reflected from the right-hand wall. The bulk of the signal in each direction, consequently,

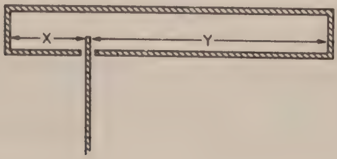


Fig. 4-2. Simple resonant cavity.

is supplied by reflection from the opposite end, and only a small amount of energy needs to be supplied by the probe. For a given signal level in the cavity, the required input power is a minimum when X and Y are odd multiples of a quarter-guide wavelength. Alternatively, for a given signal input, the signal level in the cavity is greatest when X and Y are odd multiples of $\lambda_g/4$. The wavelengths or frequencies for which this is true are said to be characteristic frequencies of the cavity, and the cavity is said to resonate at these frequencies. The relation between X, Y and λ_g is tabulated in Table 4-1 for the four characteristic frequencies having the longest guide wavelengths. Note that there are several possible positions for the probe at the higher characteristic frequencies. The longest guide wavelength is the one for which the length of the cavity is $\lambda_g/2$, and the guide wavelengths of the other characteristic frequencies are in the ratio 1, ½, ⅓, ¼, etc. Unfortunately, the relation between λ_g and λ_a is not the same for all of the characteristic frequencies, and the wavelengths in air do not satisfy this simple relation, nor are the characteristic frequencies related by the reciprocals of these ratios (1, 2, 3, 4, etc). Note that a second probe can be added to take energy out of the cavity. If, then, the input probe is excited

Table 4-1. Relation Between X, Y, and λ_g

X	Y	X+Y	λ_g
$\lambda_g/4$	$\lambda_g/4$	$\lambda_g/2$	$2(X+Y)$
$\lambda_g/4$	$3\lambda_g/4$	λ_g	$(X+Y) = 2(X+Y)/2$
$3\lambda_g/4$	$\lambda_g/4$	λ_g	$(X+Y) = 2(X+Y)/2$
$\lambda_g/4$	$5\lambda_g/4$	$3\lambda_g/2$	$2(X+Y)/3$
$3\lambda_g/4$	$3\lambda_g/4$	$3\lambda_g/2$	$2(X+Y)/3$
$5\lambda_g/4$	$\lambda_g/4$	$3\lambda_g/2$	$2(X+Y)/3$
$\lambda_g/4$	$7\lambda_g/4$	$4\lambda_g/2$	$2(X+Y)/4$
$3\lambda_g/4$	$5\lambda_g/4$	$4\lambda_g/2$	$2(X+Y)/4$
$5\lambda_g/4$	$3\lambda_g/4$	$4\lambda_g/2$	$2(X+Y)/4$
$7\lambda_g/4$	$\lambda_g/4$	$4\lambda_g/2$	$2(X+Y)/4$

at a constant signal level, the signal level in the cavity will be a maximum when the cavity resonates (the resonant frequency of the cavity may be changed by tuning it, or the frequency of the input signal may be changed to obtain resonance) and the energy transferred to the output probe also will be a maximum at resonance, since the output probe takes energy from the cavity.

PRINCIPLES OF RESONANT CIRCUITS

Since the behavior of a cavity resonator resembles that of an LC circuit as well as that of a transmission line section, it is convenient to introduce the detailed analysis of resonator behavior by examining the behavior of the two conventional circuits.

Resonant LC Circuit

Consider the shunt-resonant, zero-resistance LC circuit shown in Fig. 4-3. This circuit consists of inductor *L* in parallel with capacitor *C*. When an alternating voltage is applied to the terminals of this circuit, the phase of the current through the inductor is 90° behind the phase of the applied voltage. The phase of the current through the capacitor is 90° ahead of the phase of the applied voltage, and the two currents (through the inductor and the capacitor) therefore differ in phase by 180°. The two currents cancel each other to some extent at any frequency; however, at the resonant frequency, f_r, the two currents are equal in magnitude and opposite in phase. Consequently, they cancel each other completely and no current flows through the circuit from the external source (although heavy circulating current flows in the circuit loop). The resonant frequency is defined by the relation

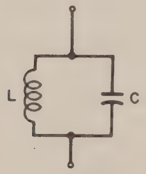


Fig. 4-3. Resonant LC circuit.

$$f_c = \frac{1}{2\pi\sqrt{LC}}$$

The impedance of this circuit is infinite at its resonant frequency—it passes no current. In practice the inductor has some resistance and the capacitor may have both lead resistance and leakage. The impedance at resonance in the practical case is not infinite, but only very great. For the purpose of discussion, it is sufficient to note that the impedance is infinite or very high at the resonant frequency.

Resonant Transmission-Line Section

The circuit shown in Fig. 4-4A is nothing more than a section of transmission line, of length *X*, and terminated by a short circuit. The previous discussion of stub tuning makes it clear that there will be a reflected signal whenever an alternating voltage is applied to the input terminals of the shorted section of transmission line. The amplitude of the reflected signal will be equal to that of the direct signal and, at the termination, their phases will differ by 180 . This condition is imposed by the fact that the terminating impedance is zero. The voltage at the termination is zero, since the direct and reflected signals have equal amplitudes and differ in phase by 180°. The current, however, is not zero, since the currents of the direct and the reflected signals add directly; forward current for the direct signal flows in the same direction as backward current for the reflected signal. At a point a quarter wavelength from the termination, the phase of the direct signal is advanced by 90°, and that of the reflected signal is retarded by 90°. The two signals, which differ in phase by 180° at the termination, are therefore in phase with each other at a point a quarter wavelength from the termination. The voltages of the two signals then add, whereas the currents cancel each other (the forward currents of the direct and reflected signals flow in opposite directions). Since the current is zero, the impedance at a point a quarter wavelength from the termination is infinite.

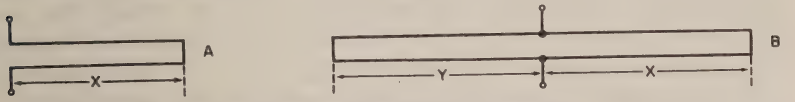


Fig. 4-4. Resonant transmission-line sections.

The same condition is obtained at all points an integral number of half wavelengths from this point (points an odd number of quarter wavelengths from the termination), since moving a half wavelength along the line changes the phase of each signal by 180° and therefore leaves them in phase with each other. In practice, of course, the impedance is very high —but not infinite. The behavior of this circuit, then, is described by the statement that the input impedance is infinite whenever the frequency of the applied signal is such that *X* is an odd number of quarter wavelengths. Now, consider the circuit shown in Fig. 4-4B, which is a section of transmission line terminated at each end by a short circuit (with input terminals connected at a suitable point). This will have an infinite input impedance if *X* is an odd number of quarter wavelengths and *Y* is also an odd number of quarter-wavelengths. The total length of the section of transmission line, for an infinite input impedance at a suitable point is, then, (X + Y), which is an even number of quarter-wavelengths (since the sum of two odd numbers is an even number) or some multiple of a half wavelength. Consequently, there will be a point (or a number of points) an odd number of quarter wavelengths from each end of the section at which the impedance is infinite; this is true whenever the frequency is such that the total length of the section is an exact multiple of a half wavelength. The characteristic wavelengths for which the input impedance is infinite are given by the relation:

$$\lambda_k = \frac{2(X + Y)}{k}$$

where, k is a positive integer (1, 2, 3, etc.), and where the wavelength is, as noted, the wavelength on the transmission line. The dominant mode of resonance is the one having the greatest wavelength. The characteristic wavelength for the dominant mode is therefore

$$\lambda_1 = \frac{2(X + Y)}{1} = 2(X + Y)$$

The wavelength for the *k* mode, is, of course λ_k, and this is 1/k times λ_1. Therefore it follows that f_k, the characteristic frequency of this mode, is approximately *k* times f_1. The characteristic frequencies are very nearly exact multiples of characteristic frequency f_1 of the dominant mode. The harmonic relation (exact multiples) is precise if the ratio between wavelength on the transmission line and wavelength in air is the same at all frequencies. The behavior of the circuit shown in Fig. 4-4B is described by stating that its input impedance is infinite at every characteristic frequency. Each characteristic frequency, f_k, is associated with a characteristic wavelength, λ_k, which is given by the relation:

$$\lambda_k = \frac{2Z}{k}$$

where, Z is the length of the transmission line section.

Resonant Cavity

It is impossible to write a simple expression for the characteristic frequencies of a general cavity resonator. If, however, the resonator consists of a waveguide section terminated at each end by a conducting wall, it is precisely equivalent to the section of transmission line described. Here, the characteristic guide wavelengths for which the input impedance of the resonator (at a suitable point) is infinite are given by the relation:

$$\lambda_{gk} = \frac{2Z}{k}$$

which may be translated as follows:

The wavelength in the guide, λ_{gk}, is equal to twice the length of the waveguide, 2Z, divided by k, which is any positive integer. Trouble arises as soon as an attempt is made to compute the wavelength in air and the corresponding frequency which is associated with a given λ_{gk}. In order to determine λ_{ak} from λ_{gk}, it is necessary to know the dimensions of the waveguide (its height and width if it is rectangular, its diameter, if it is circular) in addition to the length of the section. It is also necessary to know the mode in which the waveguide is operating. Finally, the relation between λ_a and λ_g may be the same for more than one mode and therefore there may be two distinct field patterns for a given characteristic frequency. The entire problem is discussed in detail for a rectangular resonator in the next paragraph. For the moment it is necessary only to point out that there is neither a single characteristic frequency, as in the LC circuit, nor a simple series of characteristic frequencies, as in the transmission-

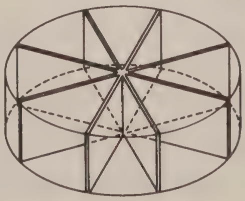


Fig. 4-5. Characteristics of cylindrical resonator.

line section. It is frequently stated that a cylindrical resonator may be thought of as an infinite number of quarter-wave shorted stubs (Fig. 4-5). This implies that the characteristic frequencies are related to each other in the same manner as those of the transmission-line section described above. This explanation of resonator behavior completely neglects the existence of various modes in waveguides and is so over simplified that it is practically of no value. The only constructive way of thinking about a cavity resonator is to consider it as a section of waveguide.

RECTANGULAR RESONATORS

Since the mathematics of rectangular cavity resonators is based on the mathematics of rectangular waveguide, it is not difficult to understand. It is possible to discuss rectangular resonators in considerable detail, using simple mathematics, whereas this is not true of other resonators. It must be kept in mind that the cavity consists of a rectangular-waveguide section terminated at each end by a conducting wall, which is assumed to be a perfect conductor.

Basic Theory

Consider a rectangular cavity resonator having a smallest dimension, height *a*, an intermediate dimension, width *b*, and a longest dimension, length *c*. The use of *c* for the length of a cavity resonator should not be confused with its use for the velocity of a light wave or a radio signal. For the moment, consider this resonator as a waveguide section of length *c*, where the waveguide of which it is composed has height *a* and width *b*. Suppose, now, that the resonator is excited at a frequency for which:

$$\lambda_g = \frac{2c}{p}$$

where,

p is a positive integer.

Because of the boundary conditions at a conducting surface, which require that the electric field be parallel to the surface and the magnetic field be perpendicular to the surface, vanish at the surface. The signal in the waveguide is reflected from the end of the cavity with a reversal of phase. The direct and incident field patterns, then, will be 180° out of phase and will, of course, be traveling in opposite directions.

At the end of the cavity, the transverse electric fields of the direct and reflected field patterns in the waveguide will cancel each other, as will the longitudinal magnetic fields. The longitudinal electric fields and the transverse magnetic fields, however, will add like the currents in the transmission line, because the forward direction for the direct signal is opposite from the forward direction for the reflected signal. At a quarter wavelength from the end of the cavity, the direct and reflected field patterns are in phase. At this point the transverse electric and longitudinal magnetic fields add, and the longitudinal electric and transverse magnetic fields cancel. At a quarter wavelength farther from the end of the cavity, the earlier situation (the phase relation that is present at the end wall) is repeated. The result is a field pattern, formed by the direct and reflected signals, which is repeated every half wavelength along the cavity. All of the distances above are measured in terms of the guide wavelength. If the waveguide is operating in either the $TE_{m,n}$ or the $TM_{m,n}$ mode, and the length of the cavity is p times $\lambda_g/2$, the resulting field pattern in the cavity will be repeated m times along the a dimension, n times along the b dimension, and p times along the c dimension.

Various Modes of Operation

In this discussion it has been assumed that the rectangular cavity was operated by making waveguide signals travel from one end of the cavity to the other. It is possible, of course, to send signals across the cavity, or between the top and bottom. The length of the waveguide section and the dimensions of the waveguide are given in Table 4-2 for a cavity of height a, width b, and length c.

Table 4-2. Waveguide Characteristics

Direction of Waveguide Signals	Length of waveguide section	Height of waveguide	Width of Waveguide
End to end	c	a	b
Side to side	b	a	c
Top to bottom	a	b	c

Waveguide Dimensions

Suppose that a rectangular cavity of height *a,* width *b,* and length *c* is excited by sending $TE_{m,n}$ signals along the cavity from end to end, at a wavelength for which λ_g is equal to 2c/p. At this point, consider what happens if the $TE_{m,n}$ signals are replaced by $TM_{m,n}$ signals having the same guide wavelength so that λ_g remains equal to 2c/p. It is assumed that *m* and *n* are the same for the TE and the TM signals. $TE_{3,5}$ signals, for example, are replaced by $TM_{3,5}$ signals. As explained in Chapter 3, a waveguide pattern having a guide wavelength:

$$\lambda_g = \frac{\lambda_a}{\cos A \ \cos B}$$

can be built up from four sets of vertically polarized waves, the paths of which make angles *A* with the top and bottom of the waveguide and angles *B* with the side walls of the guide. Also, a similar pattern can be built up from horizontally polarized waves having the same λ_a. If the same values of *m* and *n* are used, both patterns will have the same guide wavelength. Although angles *A* and *B* are different for the two patterns, the product (cos *A* and cos *B*) is the same for both. Thus, if *m, n* and λ_a are the same for both patterns, λ_g also will be the same. It has been pointed out that these patterns can be combined with one phase relation to obtain the pattern of a $TM_{m,n}$ mode. It therefore follows that a $TE_{m,n}$ mode and a $TM_{m,n}$ mode have the same guide wavelength. If *m, n* and λ_a are the same for both modes, they will have the same λ_a. It is clear, then, that λ_a is unchanged when the $TE_{m,n}$ signals, having λ_g equal to 2c/p, are replaced by $TM_{m,n}$ signals having the same *m, n* and λ_g. λ_a for either case is given by the relation:

$$\lambda_a = \frac{1}{\sqrt{\frac{(m)^2}{2a} + \frac{(n)^2}{2b} + \frac{(p)^2}{2c}}}$$

Characteristic Wavelengths

The characteristic wavelength (the value of λ_a) was calculated for a cavity of height *a,* width *b,* and length *c,* excited by $TE_{m,n}$ or $TM_{m,n}$ signals traveling along the cavity from end to end, and having a wavelength for which λ_g is equal to 2c/p. This mode of exciting the cavity produces a field pattern that is repeated *m* times along the *a* dimension, *n* times along the b dimension, and p times along the c dimension. An examination of the expression for λ_a shows that it involves a, b, c, m, n, and p, but makes no mention of the direction in which the waveguide signal travels. It follows that λ_a will have the same value (in a given cavity) for every mode of operation in which the field pattern is repeated m, n, and p times along the a, b, and c dimensions, respectively. The characteristic wavelengths are, then, those for which:

$$\lambda_a = \lambda_{m,n,p} = \frac{1}{\sqrt{\dfrac{(m)^2}{2a} + \dfrac{(n)^2}{2b} + \dfrac{(p)^2}{2c}}}$$

where, there may be more than one mode of operation of the cavity at a given characteristic wavelength. The longest characteristic wavelength is, to some extent, equivalent to the cutoff wavelength of a waveguide, since the cavity cannot resonate at longer wavelengths. The parallel is not exact, however, because the waveguide operates at any frequency above cutoff, whereas the cavity resonates only at its characteristic frequencies.

Terminology of Modes

Modes in a rectangular cavity resonator are described as TE modes, if they are excited by TE waveguide signals, and as TM modes, if they are excited by TM waveguide signals. Each mode is described further by three subscripts, which show how many times the field pattern is repeated along the height, width, and length of the cavity. There are, therefore, $TE_{m,n,p}$ modes and $TM_{m,n,p}$ modes. If the electric field is perpendicular to, for example, the top and bottom of the cavity, it is possible to have a pattern that is uniform along one dimension (in this case, the height) so that one subscript may be zero. Here, m is zero. A field that is perpendicular to two walls is, however, parallel to the other four. A pattern that is uniform in one dimension must therefore vary along both of the other dimensions in order that the electric field may satisfy the boundary conditions by vanishing at the surface of each of the four walls to which it is parallel. It is impossible, on this account, to have a mode in which the field pattern is uniform along more than one direction, and there are no modes for which more than one subscript is zero.

Field patterns that are uniform along one dimension are excited by waveguide signals like those of the $TE_{o,n}$ modes or the $TE_{m,o}$ modes in a rectangular waveguide and are, accordingly, TE modes. The cavity resonator therefore may operate in either a $TE_{m,n,p}$ or a $TM_{m,n,p}$ mode at a characteristic wavelength, $\lambda_{m,n,p}$, for which all of the subscripts are nonzero. It may operate only in a $TE_{m,n,p}$ mode at a characteristic wavelength, $\lambda_{m,n,p}$, for which one of the subscripts is zero. There are no characteristic wavelengths for which more than one subscript is zero.

Field Patterns of Modes

The dominant characteristic wavelength—the longest one—is $\lambda_{0,1,1}$. The corresponding mode is the $TE_{0,1,1}$ mode, which has the field pattern shown in Fig. 4-6A. This mode may be excited by a $TE_{0,1}$ signal traveling along the resonator from end to end, or by a $TE_{0,1}$ signal traveling across the resonator from side to side. The field pattern is the same in both instances so that there is, in fact, only a single $TE_{0,1,1}$ mode. Similar patterns are obtained in the $TE_{1,0,1}$ mode, at $\lambda_{0,1,1}$, where the electric field is perpendicular to the sides of the cavity and in the $TE_{1,0,1}$ mode, at $\lambda_{1,1,0}$, where the electric field is perpendicular to the ends of the cavity. (In the $TE_{0,1,1}$ mode the electric field is perpendicular to the top and bottom of the cavity.) All TE modes, having one subscript equal to zero, have field patterns that are repetitions of the simple $TE_{0,1,1}$ pattern. A single example, the pattern of the $TE_{0,1,2}$ mode, is shown in Fig. 4-6B.

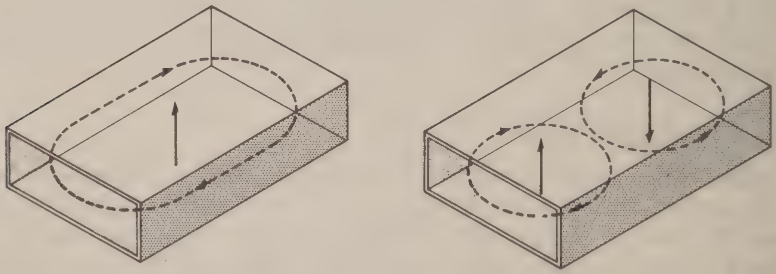

(A) $TE_{0,1,1}$ (B) $TE_{0,1,2}$

Fig. 4-6. Simple resonator modes.

The $TE_{0,1,2}$ mode may be excited by a $TE_{0,1}$ signal traveling along the resonator or by a $TE_{0,2}$ signal traveling across the resonator. The pattern, however, is the same in each case. For any characteristic wavelength having one subscript equal to zero, there is a single TE mode at which the cavity may operate. There are several possible modes, however, for a characteristic wavelength having three nonzero subscripts. At $\lambda_{1,1,1}$ for example, the cavity may be excited by a $TE_{1,1}$ signal traveling along the cavity from end to end. This developes a $TE_{1,1,1}$ mode, with the pattern shown in Fig. 4-7A. Note that the electric field is parallel to the ends of the cavity. A second $\lambda_{1,1,1}$ mode may be developed by a $TE_{1,1}$ signal traveling across the cavity. This will have a similar pattern, but the electrical field will be parallel to the sides of the cavity.

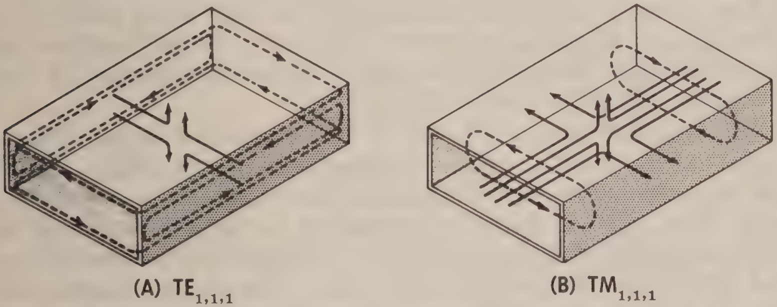

(A) $TE_{1,1,1}$ (B) $TM_{1,1,1}$

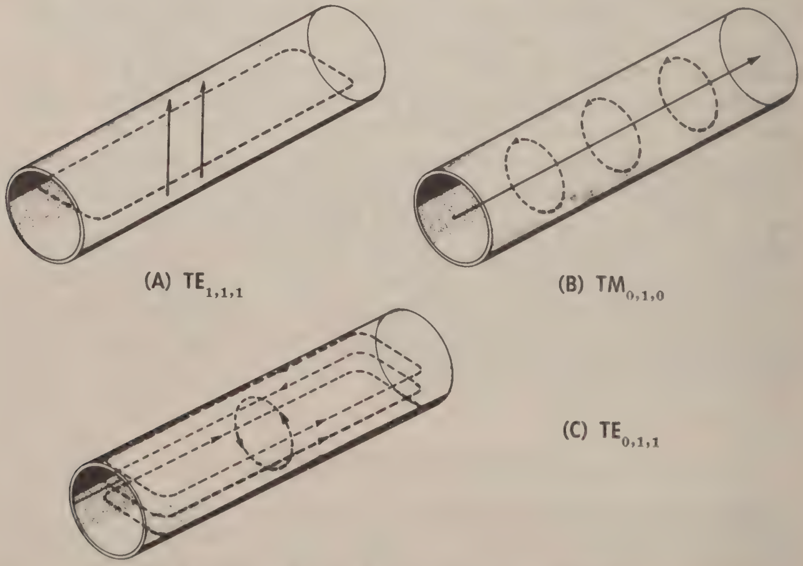
Fig. 4-7. Higher resonator modes.

Finally, there is a third $TE_{1,1,1}$ signal traveling from the top of the cavity to the bottom. Here, the electric field will be parallel to the top and bottom of the cavity. There are also, for the same characteristic wavelength, three $TM_{1,1,1}$ modes, all excited by $TM_{1,1}$ signals. In one (Fig. 4-7B) the signals travel along the cavity, from end to end, and the magnetic field is parallel to the ends of the cavity. In the other two modes, the signal travels across the cavity, with the magnetic field parallel to the sides of the cavity, or it travels from top to bottom, with the magnetic field parallel to the top and bottom of the cavity. The various patterns for characteristic wavelengths with subscripts greater than 1 are repetitions of the simple patterns. For any characteristic wavelength having three nonzero subscripts, there are three TE modes and three TM modes in which the cavity may operate.

CYLINDRICAL RESONATORS

Like a discussion of circular waveguide, a detailed discussion of cylindrical resonators involves extremely complicated mathematics beyond the scope of this book. Therefore the subject is discussed only briefly. As with a rectangular resonator, the characteristic wavelengths of a cylindrical resonator are distinguished from each other by three subscripts, m, n, and p. There is, unfortunately, no simple expression that relates $\lambda_{m,n,p}$ to the dimensions of the resonator. The subscripts m and n, like those used to describe a mode in circular waveguide, refer to the number of times the field pattern is repeated around the circumference of the cylinder, m, and the number of times that the pattern is repeated from the axis of the cylinder to its circumference, n. The third subscript, p, refers, as in the rectangular resonator, to the number of times the field pattern is repeated along the length of the resonator, along the axis of the cylinder.

The simplest mode in a cylindrical resonator is the $TE_{1,1,1}$ modes (Fig. 4-8A), which corresponds to the $TE_{0,1,1}$ mode in a rectangular resonator. Another simple mode is the $TM_{0,1,0}$ mode (Fig. 4-8B), which corresponds to the $TE_{1,1,0}$ mode in a rectangular resonator, and is seen frequently in resonators that have a modified cylindrical form. Among such resonators are the cavity of an X-band klystron, mentioned in the following paragraph and described in detail in Chapter 5, and the cavities used for TR and ATR switches. Another important mode is the $TE_{0,1,1}$ mode shown in Fig. 4-8C. A cavity resonator operated in this mode has extremely low


(A) $TE_{1,1,1}$

(B) $TM_{0,1,0}$

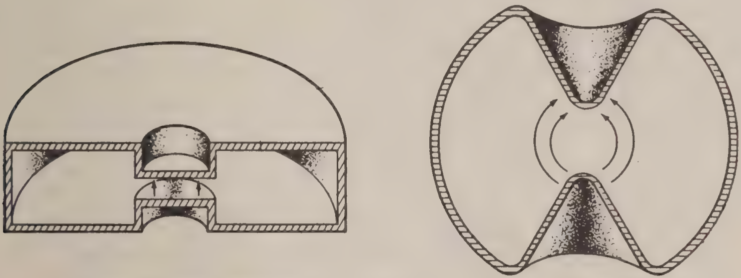
(C) $TE_{0,1,1}$

Fig. 4-8. Field patterns in cylindrical resonators.

losses. Its input impedance at resonance is therefore extremely high and the variation in impedance with change in frequency is very sharp. Microwave wavemeters are often cylindrical cavities operating in this mode.

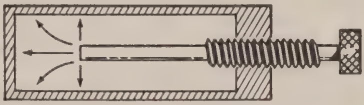
OTHER RESONATORS

Cavity resonators are seldom simply rectangular or cylindrical. The rectangular resonator has been discussed in detail to illustrate the basic theory behind all cavity resonators, and the cylindrical resonator has been described briefly because most actual resonant cavities operate in modes similar to the $TM_{0,1,0}$ and $TE_{0,1,1}$ modes of a cylindrical resonator. Three representative *re-entrant* cavities are illustrated in Fig. 4-9. The term *re-entrant* indicates that a portion of the wall protrudes into (or enters) the space within the cavity. Fig. 4-9A shows the resonant cavity of an X-band reflex klystron; Fig. 4-9B shows a TR cavity; and Fig. 4-9C represents a re-entrant wavemeter cavity. All three operate in modes similar to the $TM_{0,1,0}$ mode of a cylindrical resonator. In each, the direction of the electric field is indicated.



(A) X-band reflex-klystron resonant cavity.

(B) TR cavity.

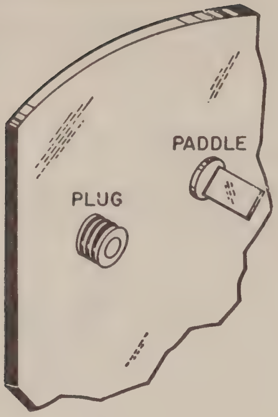


(C) Re-entrant wavemeter cavity.

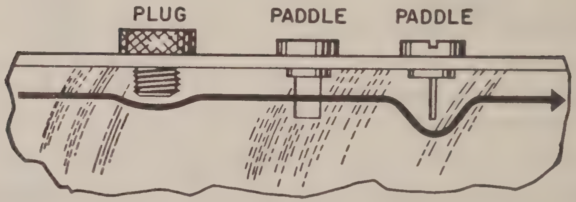
Fig. 4-9. Cavity resonators of various shapes.

CAVITY RESONATOR TUNING

One method of tuning a cavity resonator is to change its dimensions by moving one of its walls in much the same way that the termination of a tuning stub is moved. This actually



(A) Physical view.



(B) Schematic view.

Fig. 4-10. Tuning plug and paddle.

is done in some wavemeter cavities. Another method is to distort the cavity by pressure at appropriate points. A third method is to distort the electric or magnetic field, thus changing the effective dimensions of the cavity. The wavemeter cavity shown in Fig. 4-9C is tuned in this manner. Changing the position of the pin alters the electric field. The magnetic field in a cavity usually is changed by inserting plugs which extend a short distance from the wall. For large changes in the magnetic field, tuning paddles are frequently used, instead of plugs.

A tuning plug and a tuning paddle are shown in the wall of a cavity in Fig. 4-10A. Fig. 4-10B shows the consequent distortion of the magnetic field. The operation of the plug is obvious. The operation of the paddle also is simple: So long as the plane of the paddle is parallel to the direction of the magnetic field, it has no effect; when the plane of the paddle is perpendicular to the direction of the field, the distortion is greatest.

COUPLING TO CAVITY RESONATORS

Energy may be fed into or taken from a cavity resonator in three ways—by using a probe, a loop, or a slit.

Probe Coupling

Probe coupling seldom is used with resonant cavities. If the probe extends deeply into the cavity (like the probe used to feed a waveguide), it is likely to load the cavity so heavily that the resonance peak is neither very high nor very sharp. If, on the other hand, the probe extends only a small distance into the cavity, there is usually some difficulty in exciting the proper mode. For example, Fig. 4-11A shows the configuration of the electric field in a rectangular cavity resonating in the $TE_{0,1,1}$ mode, and Fig. 4-11B shows the configuration of the electric field set up by a probe. Because of this, the use of a probe is usually limited to applications where tight coupling is desired, such as to and from a waveguide, or from a directional coupler. Also, it is used where the signal is set up by some other method and only a small amount of energy is to be taken out, such as from slotted sections of waveguide or coaxial line.

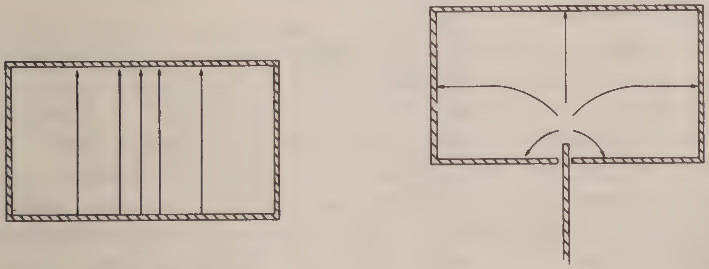

(A) Desired electric field. (B) Electric field set up by probe.

Fig. 4-11. Effect of probe coupling on electric field.

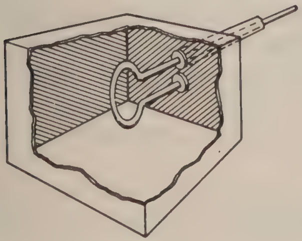


Fig. 4-12. Loop coupling to cavity resonator.

Loop Coupling

Loop coupling is more common than any other method. Fig. 4-12 shows a coupling loop used to excite the $TE_{0,1,1}$ mode in a rectangular resonator. The basic advantage of loop coupling lies in the fact that the loop does not deform the electric field pattern to the same extent that the probe does. The degree of coupling is controlled by the size of the loop and sometimes, but less frequently, by the placement and orientation of the loop. If the loop is turned so that it lies in a plane parallel to the magnetic field of a given mode, it is not coupled at all to that mode, although there may be undesired coupling to other modes. Its operation in this respect is similar to that of the tuning paddle. One of the major uses of loop coupling is to take energy from the resonant cavity of a magnetron or a klystron, and it is used also to couple to and from the resonant cavities of some wavemeters.

Slit Coupling

Slit coupling, which may involve small holes instead of slits, is used mostly in TR and ATR switches. Energy leaks into or out of the resonant cavity by means of the slits or holes in the same manner that it leaks from the main waveguide through two small holes into the auxiliary cavity (which is *not* resonant) of the directional coupler discussed previously. The principal merits of slit coupling lie in its mechanical simplicity and, in the case of TR and ATR switches, in the relatively smooth transition between the waveguide and the cavity when the coupling is tight.

TYPES OF CAVITY RESONATORS

Wavemeters

Most microwave wavemeters are essentially similar. A resonant cavity is loosely coupled to the signal source and a detector is coupled to the cavity. The cavity is then adjusted to resonate at the operating frequency, resonance being indicated by a maximum reading at the detector, and the frequency is read from a scale on the tuning adjustment. Fig. 4-13 shows the equivalent circuit of such a wavemeter; Fig. 4-14 shows a cutaway view of the tunable cavity with its input and output coupling loops.

The impedance of a resonant circuit, whether it is a cavity resonator, a section of transmission line, or an LC circuit,

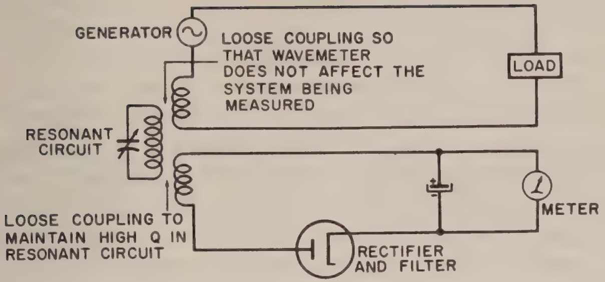

GENERATOR
LOOSE COUPLING SO THAT WAVEMETER DOES NOT AFFECT THE SYSTEM BEING MEASURED
LOAD
RESONANT CIRCUIT
METER
LOOSE COUPLING TO MAINTAIN HIGH Q IN RESONANT CIRCUIT
RECTIFIER AND FILTER

Fig. 4-13. Wavemeter equivalent circuit.

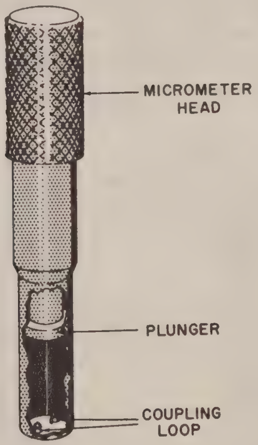

MICROMETER HEAD
PLUNGER
COUPLING LOOP

Fig. 4-14. Wavemeter tunable cavity.

reaches a maximum at its resonant frequency (or at each of its charateristic frequencies) and falls off at frequencies above and below this. As the losses in the system are increased, the peak becomes lower and less sharp. It is customary to describe the sharpness of the resonance peak by means of a quantity, *Q*, which is defined as follows:

Let Z_r be the impedance at the resonant frequency (or the characteristic frequency), f_r. At some frequency above f_r the impedance will have a magnitude equal to $Z_r \sqrt{1/2}$ or $.707Z_r$. Let this frequency be f_1. At some frequency below f_r the impedance will have the same magnitude. Let this frequency be f_2. Then, by definition,

$$Q = \frac{f_r}{(f_1 - f_2)}$$

The *Q* of a cavity resonator is commonly several thousand.

Equivalent Circuit—In Fig. 4-13 the tuned circuit is loosely coupled to the transmission line in order to avoid disturbing the system, the frequency of which is being measured. In addition, the detector is coupled very loosely to the resonant circuit in order to avoid lowering the Q. The detector shown consists of a conventional rectifier and filter circuit which feeds a meter, but, at microwave frequencies, the diode rectifier is replaced by a crystal rectifier and the filter capacitor is part of the capacitance of the crystal holder. In some cases an amplifier may be inserted between the rectifier output and the meter. The resonant circuit is adjusted so that the rectifier output is at maximum. When several frequencies are present, the meter will show a maximum as the resonant circuit is tuned through each one.

Tunable Cavity—The tunable cavity shown in Fig. 4-14 consists of a cylindrical cavity of adjustable length operating in the $TE_{0,1,1}$ mode. The upper stop is moved by a conventional micrometer head until resonance is reached. The position of the plunger is read on a scale at the micrometer head and the frequency is determined from a calibration curve. A cavity of this sort is a precision instrument and must be treated with great care.

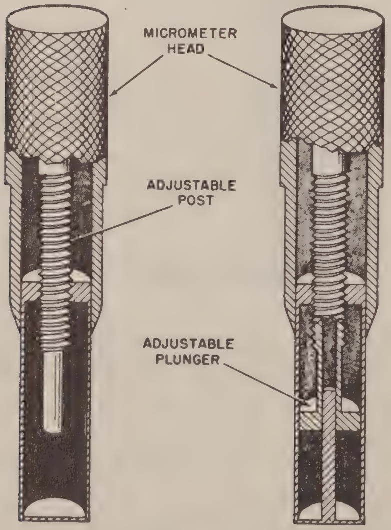


Fig. 4-15. Wavemeter tunable cavities for low microwave frequencies.

Other Forms of Cavities Used In Wavemeter—At the lower microwave frequencies, a simple cylindrical cavity becomes unduly large and various modified forms are used. Two of these types are shown in Fig. 4-15. Various forms of coupling are used, loop coupling being the most common.

Absorption Wavemeter

Another method of using a resonant cavity to determine frequency is shown in Fig. 4-16. Here, a resonant cavity is coupled tightly to a waveguide so that it absorbs a substantial amount of energy when it is tuned to resonance. As the cavity is tuned to the operating frequency, the signal in the waveguide will drop.

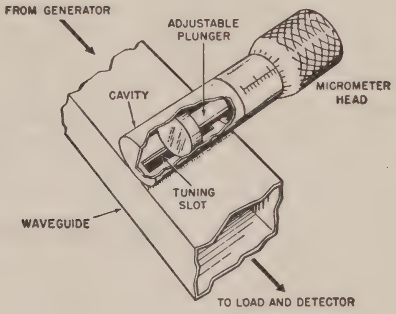


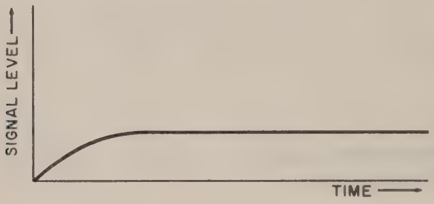
Fig. 4-16. Resonant cavity used in absorption wavemeter.

Echo Boxes

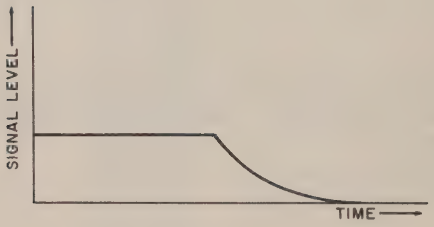
An echo box is a piece of test equipment used to check the overall performance of a radar system. It consists of a tunable resonant cavity that is connected through a transmission system to a small antenna. The cavity is usually cylindrical, of adjustable length for tuning, and operated in the $TE_{0,1,1}$ mode in order to obtain the highest possible Q.

An S-band echo box uses a small dipole antenna which is connected through flexible coaxial cable to the resonant cavity. Here, a small loop is used for coupling between the cable and the cavity. An X-band echo box generally uses a small dipole antenna at the center of a parabolic reflector. The cavity is mounted on the back of the reflector, and a short length of rigid coaxial line carries energy between the dipole and the coupling loop in the cavity. In most cases, the resonant cavity is the tuning knob in a micrometer head, and the echo box is provided with a calibration chart so that it can be used as a wavemeter. Occasionally a second coupling device is added to the cavity. The coupling device generally is a probe which is part of a crystal-holder assembly similar to those discussed in Chapter 3. The crystal rectifier output is fed to a meter so that the signal level in the cavity resonator can be measured.

If a cavity resonator is excited by an external signal, the level of the signal in the cavity will rise as shown in Fig. 4-17A. Similarly, if the excitation is stopped suddenly, the signal level in the cavity will decay as in Fig. 4-17B. It is assumed that the exciting signal is at the resonant frequency of the cavity. The decay time is related to the Q of the resonator, and the signal level will fall to .01 of its original value in 1.5Q cycles. Similarly, the rise time is related to Q. The signal level rises to .99 of its ultimate value (or reaches a point 1% short of its ultimate value) in 1.5Q cycles. The rise and decay times also can be expressed in term of the cavity time constant, which is the time during which the signal level decays to about one-third (actually .368) of its original value, or during which it rises to about two-thirds (actually .632) of its ultimate value. In practice, the time constant of the cavity is so long that the signal level cannot rise to its ultimate value during the transmitter pulse. A graph of signal level against time is shown in Fig. 4-18.



(A) Signal level versus rise time.



(B) Signal level versus decay time.

Fig. 4-17. Rise and decay time of signal in resonator.

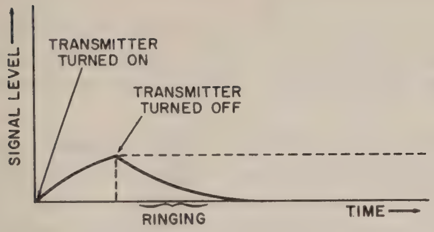


Fig. 4-18. Signal level in echo box.

Appearance of the Echo Box Signal—Fig. 4-19 shows the appearance of the echo box signal on an *A-scope*, which displays receiver output versus range or time. During the transmitter pulse, the receiver is saturated and the output signal reaches its maximum level. Following the transmitter pulse, the echo box (assuming that there are no short-range echoes). the receiver output results only from the signal picked up from The output drops at first because the receiver sensitivity has been somewhat reduced as a result of saturation by the transmitter pulse, and also because the TR (transmit-receive) and ATR (anti-transmit-receive) switches take some time to recover. The receiver, however, quickly recovers its normal sensitivity and the signal from the echo box is then sufficient to drive it to full output.

After a short interval, during which the receiver is saturated, the echo-box signal drops off so far that the receiver output also drops. From this point, the signal from the echo box and the output from the receiver both decay until the receiver output is lost in the *grass* or electric noise, on the A-scope. The *ringing time* is, of course, the time interval between the

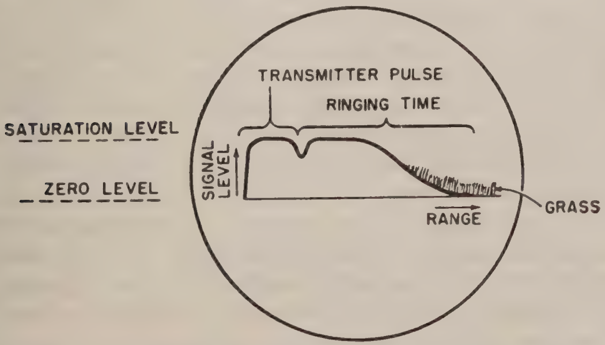
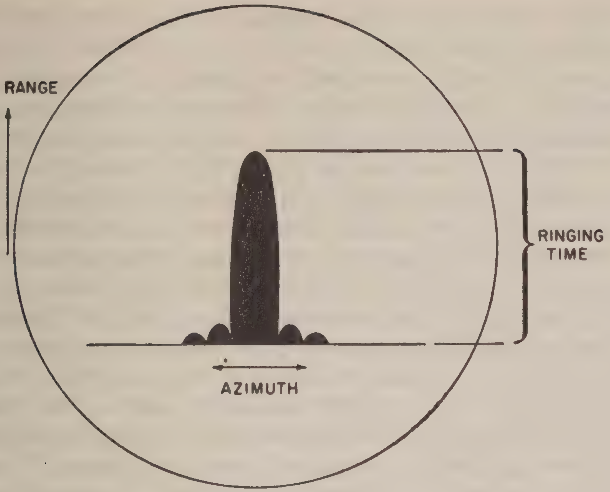


Fig. 4-19. Echo-box signal waveform displayed on an A-scope.

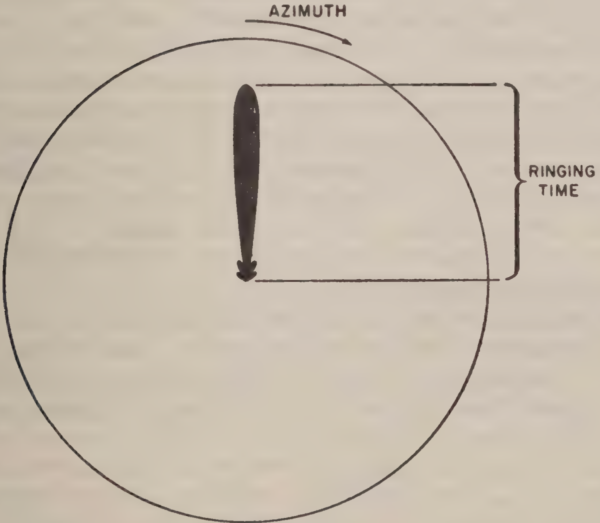
end of the transmitter pulse and the point where the echo box signal is lost in the grass. It is customary to express the ringing time in whatever range units (yards or miles) are shown in the A-scope. The ringing time may be increased in three ways:

1. Raising the transmitter output causes the echo box resonator to be excited at a higher level; it then takes longer for the echo box signal to decay to a point where the receiver output is lost in the grass.
2. Increasing the duration of the transmitter pulse has the same effect, since the signal level in the resonator will rise further toward its ultimate value if the period of excitation is short compared to the time constant of the resonator. As before, a higher initial signal level will increase the time required for the signal to decay to the point where the receiver output is lost.
3. An increase in receiver sensitivity will allow the receiver to *see* smaller signals, and, consequently, the period of decay will be longer to the lower critical level where the receiver output is lost.

In this connection, it is necessary to distinquish between receiver sensitivity and receiver gain. Increasing the sensitivity raises the ratio of signal output to noise output for a given level of the input signal. Increasing the gain, however, raises both the signal and noise output without affecting their ratio. This will affect the appearance of the picture on the A-scope, but will not alter the ringing time. The appearance of the echo box signals on a *B-scope,* which displays range versus azimuth, and on a *PPI scope* (plan-position-indication oscilloscope), which displays range versus direction, are shown in Figs. 4-20A and B, respectively. The receiver output controls the brightness of the spot on the scope in each case. The ringing time is always taken from the greatest peak, which is obtained when the antenna is pointing at the echo box. When the antenna is pointed in any other direction, the ringing time decreases because the transmitter delivers a weaker signal to the echo box, and also because the receiver output is reduced for a given signal *from* the echo box. The small peaks in the ringing-time pattern result from side lobes in the antenna pattern. Although most of the energy is radiated in the main beam, a portion is radiated in directions near the beam. The behavior of the antenna for receiving is much the same. The strongest response from the receiver is obtained when the signal source lies in the main beam of the antenna, but signals can be picked up weakly from directions close to the beam.



(A) B-scope display.



(B) PPI display.

Fig. 4-20. Echo-box signals displayed on B-scope and PPI.

This is discussed more fully in Chapter 7, which deals with microwave antennas.

Interpretation of Echo Box Test Results—In order to interpret the results of an echo box test, it is desirable to recall the expression developed for a given radar installation. Operating against a given target, the maximum range (unless limited by the horizon) is proportional to the fourth root of

P/p′, where *P* is the power output of the transmitter, and *p′* is the power input required by the receiver to develop a usable receiver output—an output signal that can be observed in spite of the electrical noise. This shows that an increase in the transmitter output, or an increase in the receiver sensitivity (which is a decrease in p′), will increase the maximum range against a given target and also will increase the ringing time with a given echo box. Therefore any adjustment of the radar system which increases the ringing time, except an increase in the duration of the transmitter pulse, will also improve the system performance. The extent of the improvement can be calculated, either from the time constant of the resonator or from its *Q*, but the process is rather complicated and the calculation is not particularly useful. However, it is possible to estimate the improvement in performance when the echo box signal is displayed on an A-scope. First, note the range at which the echo box signal is just as strong as the noise signal on the scope. Then, after making an adjustment which increases the ringing time, note the new level of the echo box signal at the range where it was previously equal

The percentage increase in range against targets of a given type will be about half of the percentage increase in the ratio of echo box signal to noise signal. The initial ratio between the two signals may be considered 1, since the point of observation is chosen as the point where the signals are equal. If, after improvement, the echo box signal is 10% greater than the noise signal, the range on targets of a given type will be improved by about 5%.

Although the echo box is one of the most valuable pieces of test equipment for tuning up a radar installation, it is dangerous to place much faith in comparisons between different ringing-time tests, unless the echo box is always placed in exactly the same position with the same orientation. Even then, any change in the temperature of the cavity resonator will alter its *Q*, and therefore will change the ringing time, as will any change in the duration of the transmitted pulse. The system should be considered satisfactory if the ringing time is within about 20% of its usual value. The echo box should then be used as an aid for the precise adjustment of the various tuning devices.

Procedure for Ringing Time Test

In view of the various types of echo boxes, the procedure is more complicated to describe than to carry out. In fact, the ringing-time test is particularly useful just because it *is* so easy. In any case, the transmitter is adjusted first. This problem is discussed in detail in Chapter 5, and the reason for mentioning it here is to point out that the proper adjustment of the transmitter insures adequate output from the magnetron at the proper pulse duration.

If the echo box is equipped with a detector, the next step is to tune it to the transmitter (by adjusting for maximum output from the detector). At this point, the transmitter should be readjusted, if necessary, so that it operates on the proper frequency. (Unless the transmitter uses a tunable magnetron, its frequency cannot be altered and the frequency check is merely one step in the check for proper overall operation.) The receiver is then tuned to the echo box signal, with the automatic-frequency-control system turned off, and the various tuning devices in the transmission system are adjusted to obtain the longest ringing time. If the echo box is not equipped with a detector, the transmitter is adjusted first and then, by observing a local echo, the receiver is tuned to the transmitter, with the automatic-frequency-control system turned off. The echo box then is tuned to the transmitter by observing the ringing time and tuning the echo box for the longest ringing time. Thereafter, the tuning devices in the transmission system are adjusted to increase the ringing time.

TR and ATR Switches

In many early radar installations the transmitter was connected to one antenna and the receiver was connected to another. Most modern radars, however, use the same antenna for transmission and reception. This eliminates the problem of pointing two antennas in the same direction and also effects

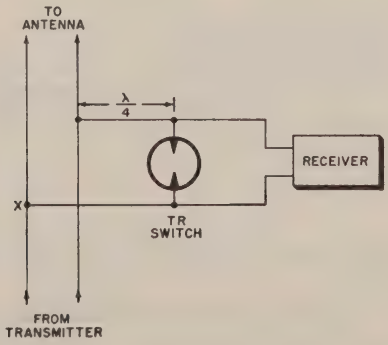


Fig. 4-21. TR switch circuit.

a reduction in the size and weight of the antenna system. This improvement introduces two problems: First, it is necessary to protect the receiver from being damaged by the high-power transmitted signal; second, it is necessary to insure that none of the incoming signal goes to the transmitter, since diversion of any incoming energy from the receiver will lower the effective receiver sensitivity. The devices used to solve these problems are called TR and ATR switches. Essentially, they are nothing more than spark gaps, but at microwave frequencies they normally involve resonant cavities.

The TR switches protect the receiver from damage by the high-power output of the transmitter (Fig. 4-21) and insure that the bulk of the transmitter output goes to the antenna. Without the TR switch, a portion of the transmitter output would go directly to the receiver instead of being radiated from the antenna. When the transmitter is on, the signal level in the line is sufficiently high to cause the TR switch to arc and short-circuit the line to the receiver. The line to the receiver, when viewed from *X* in Fig. 4-21, then appears as a quarter-wave shorted stub and has an infinite impedance. All of the transmitter output reaches the antenna except for the small amount necessary to maintain the arc across the TR switch. In addition, only a small amount of energy reaches the receiver, which this is protected from damage. As soon as the transmitter is turned off, the arc goes out and the receiver is automatically reconnected to the antenna. With very high-powered transmitters, two spark gaps sometimes are used to afford added protection to the receiver. The spark gap nearer the transmission line is called a pre-TR switch and the one nearer the receiver is referred to as the TR switch (Fig. 4-22).

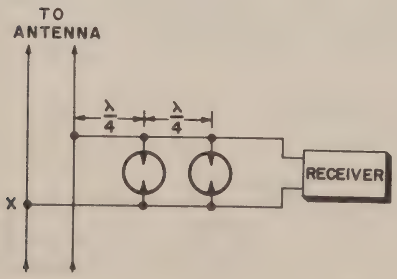


Fig. 4-22. Pre-TR switch TR switch circuit.

The TR switch protects the receiver from damage and also prevents diversion of any significant portion of the transmitter

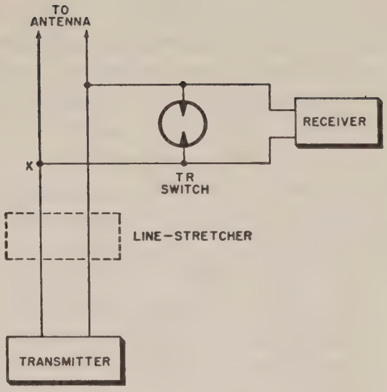

TO ANTENNA
RECEIVER
X
TR SWITCH
LINE-STRETCHER
TRANSMITTER

Fig. 4-23. Line stretcher in TR switch circuit.

output from the antenna. It is not, however, effective during reception and cannot prevent the incoming signals from being partly wasted at the transmitter. One method of preventing this waste of incoming energy (Fig. 4-23) is to use a line-stretcher between junction *X* and the transmitter. The length of the line between *X* and the transmitter is adjusted, by use of the line-stretcher, so that impedance at *X* looking toward the transmitter (with the transmitter turned off) is infinite. All of the incoming energy then goes to the receiver. Sometimes the line between *X* and the transmitter is preplumbed and the line stretcher is omitted.

A more common method of insuring that all the incoming energy reaches the receiver is by use of an ATR switch, as

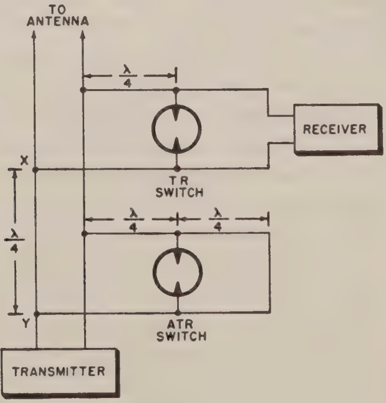

TO ANTENNA
$\frac{\lambda}{4}$
RECEIVER
X
TR SWITCH
$\frac{\lambda}{4}$
$\frac{\lambda}{4}$
$\frac{\lambda}{4}$
Y
ATR SWITCH
TRANSMITTER

Fig. 4-24. ATR switch circuit.

shown in Fig. 4-24. When the transmitter is on, both the TR and ATR switches arc. The impedance looking from X toward the TR switch is then infinite, since this branch of the line looks like a quarter-wave shorted stub. For the same reason, the impedance at Y looking toward the ATR switch is also infinite. All of the transmitter output, except that necessary to maintain the arcs, reaches the antenna, and the receiver is protected. When the transmitter is turned off, the arcs go out and the receiver is connected to the transmission line as explained above. In addition, the impedance at point Y looking toward the ATR switch is now zero, since this branch is now a half-wave shorted stub. The line then is effectively short-circuited at Y and the impedance at X looking toward the transmitter is that of a shorted quarterwave stub which is infinite, and all of the incoming energy goes to the receiver. The effect of the combined switch system is as follows: The receiver is protected from damage—

1. When the transmitter is on, the receiver is removed from the transmission line and all of the transmitter output goes to the antenna.
2. When the transmitter is off, the receiver is connected to the transmission line and the transmitter is removed from the line so that all of the incoming energy goes to the receiver.

Construction—The TR and ATR switches are essentially similar in construction. Each consists of a resonant cavity with a pair of electrodes across which an arc is struck when the transmitter is turned on. Coupling to and from the cavity is by means of loops when the switch is used with coaxial line, and by means of windows (which are large slits) when the switch is used with waveguide. The cavity is normally equipped with both tuning paddles and tuning plugs, the paddles being used for coarse tuning and the plugs for fine tuning. In order that the arc may be easily formed, the electrodes are contained in a glass bulb filled with gas under low pressure. The complete assembly (bulb with electrodes) is called a TR tube. One of the electrodes frequently is mounted on a flexible diaphragm which forms part of the bulb wall; the spark gap spacing can then be adjusted by deforming the diaphragm.

It is customary to provide a third electrode, called a *keep-alive* electrode. A voltage, called the keep-alive voltage, is applied to this electrode and a steady glow discharge is maintained between one side of the spark gap and the keep-alive electrode. This subsidiary arc permits the main arc to form more quickly when the transmitter is turned on. Also, it is necessary to provide a current-limiting resistor in the keep-alive line. This must be placed close to the TR tube to avoid oscillation of the keep-alive arc (which is similar to the oscillation of a thyratron oscillator) and consequent damage to the receiver when the TR switch fails to arc.

Chapter 5

Microwave Transmitter Oscillators

The preceding chapters have demonstrated the marked differences between microwave circuit techniques and those used at lower frequencies. A conventional transmission line, for example, is often replaced by a waveguide at microwave frequencies, and a cavity resonator takes the place of the more familiar coil-capacitor combination. In view of these differences, it is not surprising that microwave oscillators have little in common with the usual vacuum-tube oscillators used in lower frequencies.

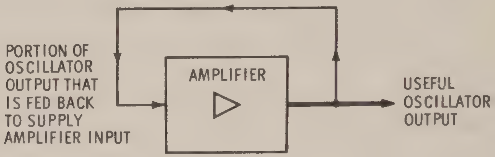
KLYSTRONS AND MAGNETRONS

At present there is no satisfactory method of amplifying microwave signals at high power levels. Microwave oscillators must therefore be connected directly to the loads they feed. Two distinct types of microwave oscillators have been developed for low-power and high-power applications. A *klystron* is used at microwave frequencies where a simple oscillator (or an oscillator and a buffer) would be used at lower frequencies. The average power output of a klystron is generally something less than 1 watt. Klystrons are used in signal generators and other test equipment, in low-power transmitters and as local oscillators in microwave superheterodyne receivers. A *magnetron*, on the other hand, is used at microwave frequencies where an oscillator and a power amplifier would be used at lower frequencies. The usual microwave transmitter with medium- or high-power output uses a magnetron oscillator. *The complete transmitter consists of the magnetron, a microwave transmission system (generally using waveguide), an antenna, a power supply, and a modulator to provide amplitude, frequency, or pulse modulation.* The magnetron thus takes the place of a conventional low-frequency oscillator, buffer, frequency-multiplier, voltage amplifier, driver, and power amplifier. Both types of microwave oscillators are vacuum tubes of a rather special sort called *velocity-modulated* tubes. The internal elements of a velocity-modulated tube act on the electron stream by changing the average velocity of the electrons while those of a conventional vacuum tube control the number of electrons in the stream.

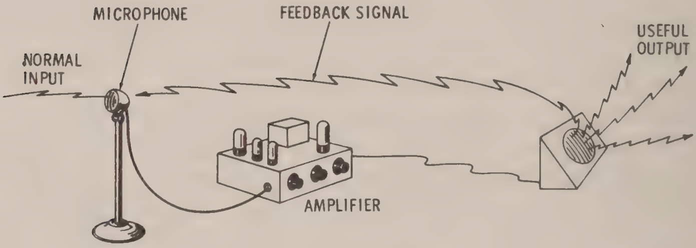
Since each type of microwave oscillator makes use of velocity modulation—a technique not ordinarily used at lower frequencies—a study of either klystrons or magnetrons must include two distinct subjects. First, how the tubes work, how they are constructed, and how the tube elements and the electron stream act on each other; and, second, why the tubes oscillate. Therefore, a general understanding is necessary, and a review of oscillator theory will make it easier to understand how klystrons and magnetrons generate microwave signals. An explanation of conventional vacuum-tube oscillators and their limitations will cover different types of vacuum tubes required for operation at microwave frequencies. The remainder of this section is, accordingly, devoted to a discussion of three basic types of oscillators—feedback, relaxation, and ringing.

FEEDBACK OSCILLATORS

Feedback oscillators, which are quite simple, are probably more common than any other type in the field of electronics. A feedback oscillator is nothing more than an amplifier that supplies its own input. Fig. 5-1A shows a block diagram of the basic feedback-amplifier loop. A small portion of the amplifier output is fed back to supply the amplifier input, and the remaining portion is delivered to an external load as the output of the oscillator. Fig. 5-1B shows a type of feedback oscillator often found when an inexperienced technician sets up a public address system. Although this is a highly undesirable system in practice, it is a good illustration of a feedback oscillator. The microphone, electronic amplifier, and speaker form an acoustic amplifier which receives an input of low-level sound and delivers an output of sound at a higher level. Here the feedback signal is in the form of sound. Following the terminology used in Fig. 5-1A, the signal delivered by the speaker is labeled *useful output,* although one familiar with the problems of public address systems may find it difficult to say what it can be used for when the system is oscillating.


PORTION OF OSCILLATOR OUTPUT THAT IS FED BACK TO SUPPLY AMPLIFIER INPUT
AMPLIFIER
USEFUL OSCILLATOR OUTPUT

(A) Basic system.


MICROPHONE
FEEDBACK SIGNAL
NORMAL INPUT
USEFUL OUTPUT
AMPLIFIER

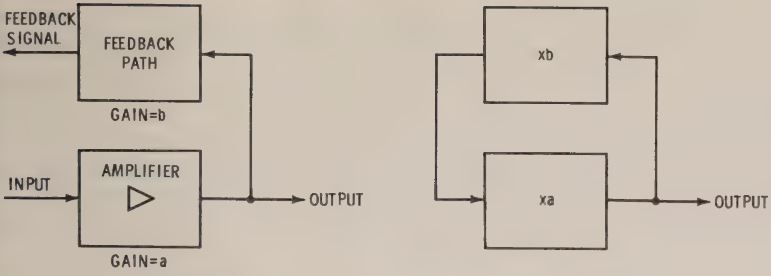
(B) PA system oscillating.

Fig. 5-1. Basic feedback-oscillator loop.

Loop Gain

The system shown in Fig. 5-2A consists of an amplifier having a gain of *a* and a feedback path having a gain of *b* (or an attenuation of 1/*b*). Since the operation of an ordinary low-power electronic amplifier is described most conveniently in terms of input and output voltage levels, it is assumed that *a* and *b* are voltage, rather than power, gains. With a system of this sort, the output is *a* times the input, and the feedback signal is *a*b times the input. If product *a*, which is the *loop gain,* is exactly equal to 1, the loop may be closed as shown in Fig. 5-2B, and the feedback signal will supply exactly the amplifier input that is required for steady oscillation.

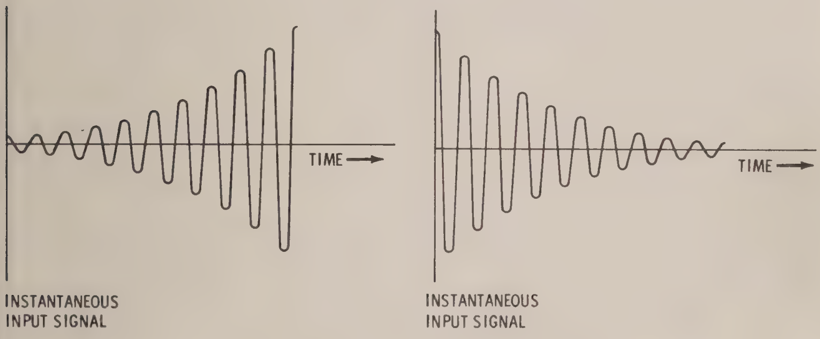
Suppose, for example, that a 1-volt signal from an external source is connected to the amplifier input in the system shown



(A) Open loop. (B) Closed loop.

Fig. 5-2. Concept of loop gain.

in Fig. 5-3A. This will develop a feedback voltage of exactly 1 volt with the same phase as the amplifier input. The feedback signal may thus be substituted for the external signal by closing the loop, as in the illustration, with no change in the signal levels at various parts of the system. If the loop gain is less than 1, however, the closed-loop system is unable to supply itself with an adequate input signal, and the oscillation will die away, as shown in Fig. 5-3B. If, on the other hand, the loop gain is greater than 1, the feedback signal is greater than the required input for steady oscillation and the amplitude of oscillation will increase so long as the loop gain exceeds 1, as in Fig. 5-3A.



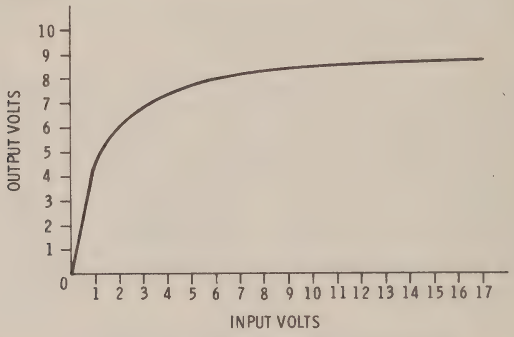
(A) Gain greater than 1. (B) Gain less than 1.

Fig. 5-3. Effect of changing loop gain.

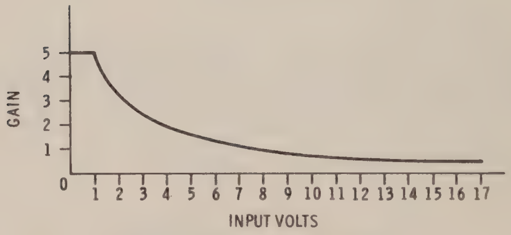
Establishing Steady Oscillation

Although this discussion explains how oscillations can sustain themselves, grow, or die away in a feedback oscillator, nothing has been said about how oscillations start or why the oscillator settles on a particular frequency. The key to this lies in *saturation* of the amplifier portion of the feedback oscillator.

Saturation, which should be familiar to every electronics technician, is the reduction in the gain of an amplifier when the signal level increases beyond some critical point. Fig. 5-4A shows the output of an amplifier for various input levels. For inputs of less than 1 volt, the output is five times the input, and the gain (Fig. 5-4B) is 5. Although the output continues to rise as the input signal is increased, the gain (the ratio of output to input) decreases. For an input of 2 volts, the output is only 6 volts and the gain is reduced to 3. By the time the input reaches 5 volts, the gain has fallen to 1.5; for an input of 8.25 volts, the gain is only 1.



(A) Output voltage versus input voltage.



(B) Gain versus input voltage.

Fig. 5-4. Amplifier-saturation graph.

With still higher input signals the output continues to rise slightly, approaching a saturation level of 8.5 volts. Whereas the gain continues to decrease, reaching 0.5 at a 17 volt input level. This set of figures has been chosen for the purpose of making a clear presentation of saturation. In practical amplifiers, however, the gain for small signals is likely to be 50 or more and may be as great as several hundred. With a clear picture of saturation in mind, it is easy to see how the oscillator behaves. With no signal in the amplifier, loop gain ab is high. Suppose, then, that there is some disturbance in the circuit—probably the result of tube noise. This disturbance may be thought of as composed of a number of distinct frequencies. For some of these frequencies, the loop gain will be high and with the proper phase for the amplifier to supply its own input. Oscillations will start and grow rapidly at all of these frequencies. As soon as the signal level in the amplifier rises, the gain (and consequently the loop gain) will be reduced. As the oscillations grow, the loop gain will become smaller until it is less than 1 for some of the frequencies of oscillations (for which the loop gain for small signals was greater than 1). The oscillations at these frequencies then die away quite rapidly. The process continues as long as the overall signal level increases (as long as the loop gain remains greater than 1 for one or more frequencies).

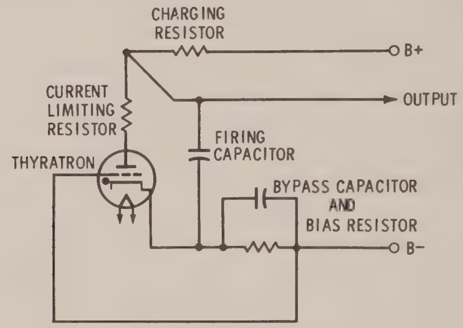
In a well designed oscillator all signals but those at one frequency die away shortly after the oscillator is turned on. With poor design, however, there may be two or more frequencies at which the oscillator operates, and in some types of oscillator, it is extremely difficult to confine the operation to a single frequency. The oscillator, therefore, reaches stable operation at a single frequency with a signal level for which the amplifier gain is $1/b$ (so that the loop gain is 1). All other frequencies are suppressed because ab is less than 1 for these frequencies. The signal level is stable, since any increase in signal level causes a reduction in loop gain and a consequent dying away of the oscillation. Similarly, any reduction in level causes an increase in loop gain and a growth of the oscillation.

RELAXATION OSCILLATORS

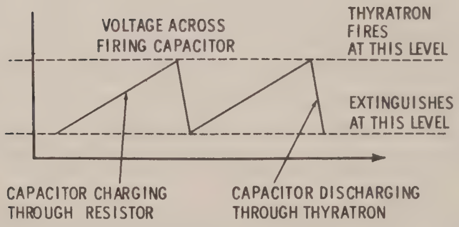
A relaxation oscillator is probably easier to understand than any other type. It consists of a system, of almost any sort, that tends to *relax* toward a preferred state. Before the preferred state is reached, a device is triggered that pushes the system *away* from the preferred state, after which the system starts once more to relax. The process is something like what would happen if a tired man were to sit on a tack, provided that for some reason he attempted to sit on the tack again and again. The period of a relaxation oscillator is the time interval between successive triggering of the displacement device. This interval depends on how far the system is displaced from its preferred state and on how rapidly it relaxes. The time required for displacement *from* the preferred state is usually negligible.

Thyratron Oscillator

The best known relaxation oscillator is undoubtedly the thyratron sweep circuit used in cathode-ray oscilloscopes of moderate speed. One version of this oscillator is shown in Fig. 5-5A. The preferred state is, of course, the condition in which the firing capacitor is fully charged so that no current flows through the charging resistor (or through the bias network). Before this state is reached, however, the thyratron fires and discharges the firing capacitor almost to zero. The capacitor then charges through the charging resistor until the voltage across it is once more sufficient to cause the thyratron to fire. The waveform of the voltage across the firing capacitor


CHARGING RESISTOR
B+
OUTPUT
CURRENT LIMITING RESISTOR
FIRING CAPACITOR
THYRATRON
BYPASS CAPACITOR AND BIAS RESISTOR
B−

(A) Circuit.

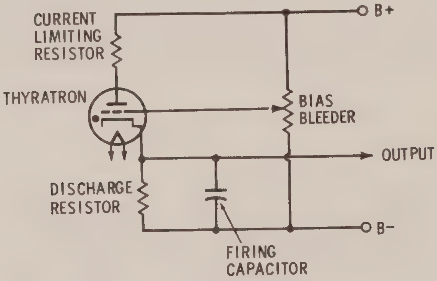


(B) Waveform.

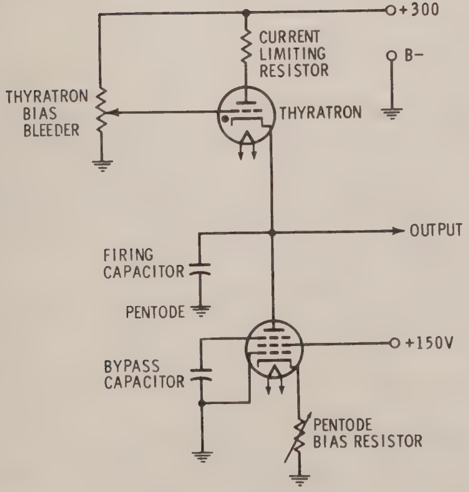
Fig. 5-5. Thyratron relaxation-oscillator circuit.

is shown in Fig. 5-5B. The charging rate—the rate at which the system approaches its preferred state—is set by the time constant of the firing capacitor and charging resistor.

An interesting variation of this circuit is shown in Fig. 5-6A. Here, the firing capacitor discharges through the discharge resistor until the thyratron cathode drops enough for the thyratron to fire. The capacitor then is recharged by the thyratron, which extinguishes itself when the anode-cathode voltage falls below a critical value. Since the thyratron fires at a critical value of grid-cathode voltage (whereas the earlier circuit fires at a critical value of anode-cathode voltage), the firing point can be adjusted easily by altering the grid bias. This, of course, changes the instant at which the system is pushed


CURRENT LIMITING RESISTOR
B+
THYRATRON
BIAS BLEEDER
OUTPUT
DISCHARGE RESISTOR
B−
FIRING CAPACITOR

(A) Simple modification.


+300
CURRENT LIMITING RESISTOR
B−
THYRATRON BIAS BLEEDER
THYRATRON
OUTPUT
FIRING CAPACITOR
PENTODE
+150V
BYPASS CAPACITOR
PENTODE BIAS RESISTOR

(B) Circuit providing more linear charging rate.

Fig. 5-6. Modified thyratron oscillator circuits.

away from its preferred state, and thus controls the frequency of oscillation. A modification of this circuit that provides a more linear charging rate is shown in Fig. 5-6B. Here, the charging resistor is replaced by a pentode that functions as a constant-current device. Coarse-frequency control is now achieved by adjustment of the thyratron bias; and fine-frequency control is obtained through adjustment of the pentode bias resistor.

Multivibrator

A conventional multivibrator (Fig. 5-7) may be thought of as a push-pull relaxation oscillator. The preferred state of each tube is the conducting state; as one tube reaches its preferred state and starts to conduct, it drives the grid of the other tube past cutoff. It is desirable to examine the manner in which the multivibrator starts to oscillate.

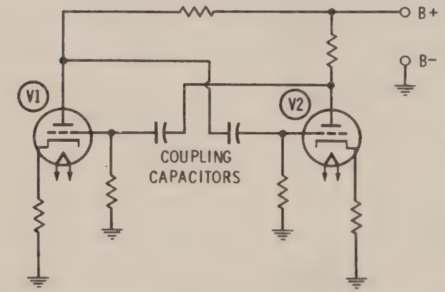


Fig. 5-7. Multivibrator circuit.

Suppose that the two tubes are in the circuit, as shown, with the single exception that the two coupling capacitors are omitted. The current through each tube then assumes a steady value which is determined by the anode and cathode resistors, and the fact that the grid is returned to B-minus. Now, if the coupling capacitors are connected, there is no change in the system at first. Any variation in the grid voltage of either tube, however, will throw the system into oscillation. Suppose that as a result of tube noise in either one of the tubes the grid of tube V1 moves slightly in the positive direction. This momentary positive signal is amplified in V1 and drives the grid of V2 in the negative direction. Tube V2, in turn, amplifies the signal at its grid and returns a large positive signal to the grid of V1. The process is cumulative and proceeds with increasing speed—the grid of tube V1 moves in the positive direction while the grid of tube V2 moves in the negative direction.

This continues until the two-tube amplification can proceed no further—usually as a result of cutoff in the tube, the grid of which is driven in the negative direction. The process stops when V2 is cut off, with the grid of V1 well above its normal level. This condition is not permanent, however, because the coupling capacitor that feeds the grid of V2 begins to shift its charge, discharging through the grid resistor of tube V2 and the anode resistor of V1. Tube V2 then starts to conduct, and the two-tube amplification process lifts the grid back into the positive region and drives the grid of V1 well past cutoff, as explained above. The behavior of the system as it goes into oscillation is a little like that of a pencil that has been balanced on end. Any disturbance will throw it out of its condition of unstable equilibrium and into oscillation. The circuit can be made to have greater amplification by connecting the cathodes of the two tubes to B-minus through a single cathode resistor —this is frequently done.

Univibrator

The circuit known as a *univibrator* relaxes toward a preferred position and does not trigger the device which displaces it. This circuit, which is sometimes called a one-shot multivibrator in the United States and a Kipp relay in Great Britain, is shown in Fig. 5-8. In the preferred state of the system, V2 conducts, since its grid is returned to its cathode and the tube has a grid bias of zero. The bias bleeder for V1 is adjusted so that V1 is just a little past cut-off. In the quiescent state, then, V2 draws current and V1 does not. The circuit is triggered by a positive pulse at the input terminal,

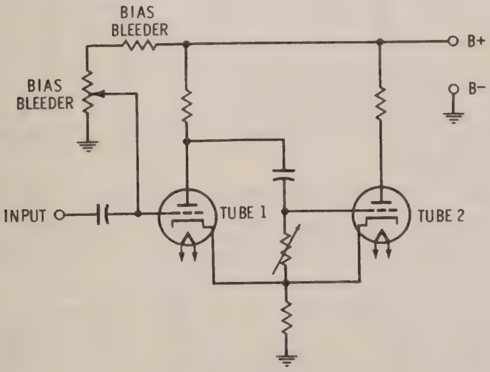


Fig. 5-8. Univibrator circuit.

and the input need be only large enough to cause tube V1 to draw current. This signal is amplified in V1 and appears as a negative pulse at the grid of V2. The resulting drop in the current of V2 causes the cathodes of both tubes to move in the negative direction, and the negative signal at the cathode of V1 causes precisely the same result as the original positive signal at the grid of V1. As in the multivibrator, the process continues until V2 is cut off; at this point, V1 draws current and V2 does not. The coupling capacitor that feeds the grid of V2 then discharges, and as soon as V2 starts to conduct, the initial process is reversed and V1 again is cut off. At this point, however, the system rests until another input pulse is received. The duration of the unstable state (with V1 conducting and V2 cut off) is determined by the resistor and capacitor connected to the grid of V2.

Relay Relaxation Oscillator

It is necessary to emphasize the fact that an oscillator need not be electronic. In the next discussion (on ringing oscillators), for instance, an old-fashioned pendulum clock is mentioned as an example.

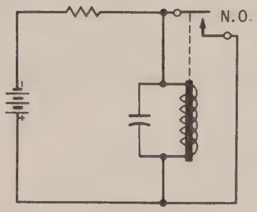


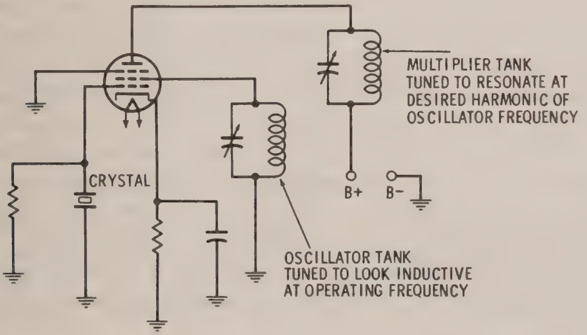
Fig. 5-9. Relay relaxation oscillator.

An electrical, but *nonelectronic* oscillator, which consists of the usual capacitor and charging resistor together with a relay, is shown in Fig. 5-9. In operation the capacitor charges through the resistor until the voltage across it is sufficient to pull in the relay. When the relay is actuated, one set of contacts is used to short the capacitor, after which the relay drops open and the capacitor starts once more toward the preferred position of full charge.

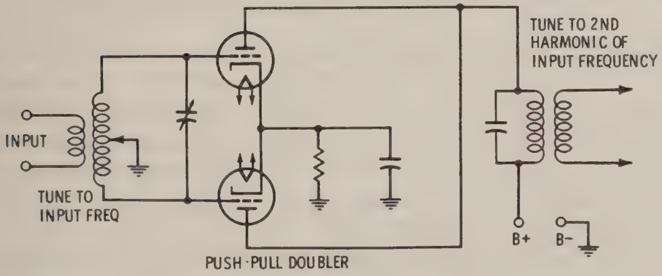
RINGING OSCILLATORS

Another type of oscillator is called a ringing oscillator. This consists of a resonant system, which may be a pendulum or a cavity resonator, together with a mechanism of some sort for feeding energy to the resonator. It has been explanied in Chapter 4 that a cavity resonator will ring at its own resonant frequency whenever it is shock-excited, and the same thing is true of any other resonator. Then if matters can be arranged so that the resonator is periodically given a small boost at the proper time, it will continue to ring and the combination of resonator and boosting mechanism legitimately may be considered an oscillator.

One good example of a ringing oscillator is the pendulum of an old-fashioned clock. Left to itself, the pendulum would swing from side to side at its resonant frequency, but the amplitude of its swings would decrease continually and finally come to rest. The clock mechanism, however, provides a gentle push to the pendulum at each end of its swing, and it goes on oscillating indefinitely so long as someone winds the clock. Another example, of course, is the balance wheel and escapement mechanism on a watch or modern nonelectric clock. This, however, may not be as familiar as the pendulum. Still


MULTIPLIER TANK
TUNED TO RESONATE AT
DESIRED HARMONIC OF
OSCILLATOR FREQUENCY
CRYSTAL
B+
B−
OSCILLATOR TANK
TUNED TO LOOK INDUCTIVE
AT OPERATING FREQUENCY

(A) Tri-tet harmonic oscillator.


TUNE TO 2ND
HARMONIC OF
INPUT FREQUENCY
INPUT
TUNE TO
INPUT FREQ
B+
B−
PUSH-PULL DOUBLER

(B) Push-pull doubler.

Fig. 5-10. Oscillator and doubler circuits.

another example of a ringing oscillator is a child in a swing. If he pumps at the right time during each swinging cycle, he will oscillate indefinitely (or until he gets hungry and goes home for supper).

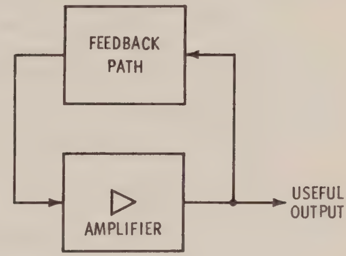
A good example of the ringing oscillator in the field of conventional electronics is the tri-tet harmonic oscillator shown in Fig. 5-10A. This consists of a pentode that operates like a combination of a triode oscillator and a buffer stage. The cathode, control grid, and screen grid of the tube are connected as a triode oscillator of one sort or another; the illustration shows them connected as a conventional crystal oscillator. The tuned circuit in the plate line is set to resonate at the desired harmonic of the oscillator frequency. Its behavior is very much like that of a clock pendulum, except that it gets a boost in each direction only once during each cycle of the oscillator. Since the resonator is boosted one way when the oscillator current is at a maximum, and the other way one-half cycle later when the oscillator current is at a minimum, the system works best as a generator of odd harmonics.

It should be noted that all frequency multipliers are not ringing oscillators. The conventional push-push doubler, for example, shown in Fig. 5-10B, operates with the inputs to the two tubes in push-pull and the two outputs in parallel. If the tubes are operated class B, the output looks like that of a full-wave rectifier and it has a very large second harmonic content. The output tank then serves as a filter rather than as a resonator (although the distinction is slight) and selects a second harmonic signal, or other even harmonic signal, which is already present. The harmonic tank in the tri-tet oscillator, however, does not receive any significant amount of harmonic signal, and therefore actually generates the harmonic output by ringing.

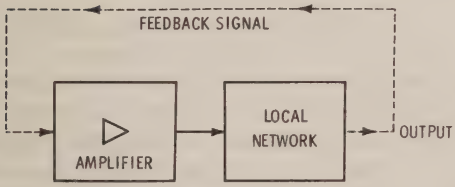
CONVENTIONAL-TUBE OSCILLATORS

The usual RF signal source at frequencies below the microwave range is a feedback oscillator of the type described previously. This part of the chapter takes up a number of representative feedback oscillator circuits and discusses some of the factors that prevent their use in the microwave region. It also includes a discussion of the Barkhausen oscillator, a type of oscillator seldom used, but mentioned here because the discussion of its operation make a good introduction to the subject of velocity-modulated tubes, which are taken up later.

The basic loop configuration shown in Fig. 5-2 is illustrated again in Fig. 5-11A, but the circuit aspect, rather than the feedback aspect, of the basic circuit is emphasized in Fig. 5-11B. Here, the useful output connection has been omitted, and the feedback path is shown as a load network fed by the amplifier. The output in Fig. 5-11B is the feedback signal, which, when the loop is closed, supplies the amplifier input. Suppose, now, that the input to the amplifier is a signal, *E*, and the output from the load network is a signal, *E'*. If there is a frequency for which *E'* and *E* are precisely in phase, with *E'* larger than *E*, the system will oscillate at that frequency when the loop is closed. It is necessary that *E'* be larger than *E*, otherwise the amplitude of oscillation will not increase. When the system is in stable operation, of course, the amplifier is partially saturated so that its gain is reduced. For signals at the normal oscillation level, *E'* and *E* are equal (they must be equal when the loop is closed, for then *E* is *E'*), but for low-level inputs with the loop open, *E'* is greater than *E*. The basic feedback amplifier therefore consists of an amplifier driving a load network that serves as the input to the amplifier. The useful output is normally taken directly from the amplifier output. Usually the amplifier consists of a single stage, either single-ended or push-pull. The requirement on the load network is that at the frequency of oscillation, the output of the load network must be slightly greater in amplitude.



(A) Basic loop.



(B) Basic circuit.

Fig. 5-11. Basic feedback-amplifier circuit.

Load Network Characteristics

A single tube can be replaced either by a voltage or current generator. In Fig. 5-12, E_{gc} is the input voltage applied between the grid and the cathode of the tube and μ,R_p, and G_m are the tube constants. The two circuits, A and B, are equivalent, but it is probably desirable to concentrate on the voltage generator circuit, since it is a more familiar one than the current generator. In this circuit the load network is fed from a resistive source having an EMF (electromotive force) proportional to $-E_{gc}$. The problem of gain may be neglected for the moment, since it is usually easy to obtain the required gain of more than 1. The load network then is required to develop an output in phase with, and of slightly greater amplitude than E_{gc} when fed from a resistive source having an EMF proportional to $-E_{gc}$. In other words, the load network, when fed through generator impedance R_p, must develop a 180° phase shift at the frequency of oscillation. This requirement is illustrated in Fig. 5-13.

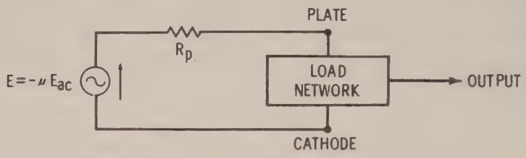


PLATE
Rp
E=-μ Eac
LOAD NETWORK
OUTPUT
CATHODE

(A) Series R_p.

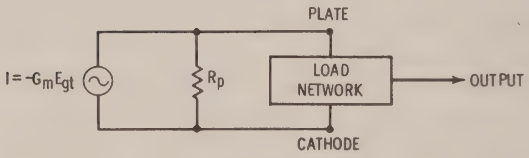


PLATE
I=-GmEgt
Rp
LOAD NETWORK
OUTPUT
CATHODE

(B) Shunt R_p.

Fig. 5-12. Equivalent amplifier circuits.

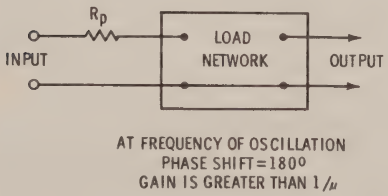
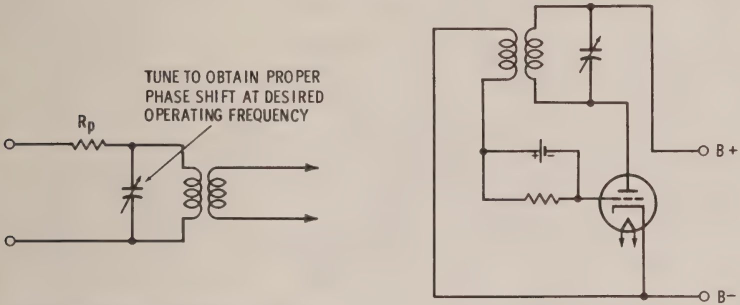

Rp
INPUT
LOAD NETWORK
OUTPUT
AT FREQUENCY OF OSCILLATION
PHASE SHIFT=180°
GAIN IS GREATER THAN 1/μ

Fig. 5-13. Load-network characteristics.

TRANSFORMER-COUPLED OSCILLATOR

One of the simplest load networks that will supply the required phase shifts of 180° is a transformer. The load network (Fig. 5-14A with the plate resistance R_p) is, then, nothing more than a simple air-core transformer with the primary tuned so that it appears resistive and supplies precisely the desired phase shift. Fig. 5-14B shows the circuit of an oscillator of this type with grid-leak bias. The parallel resistor (the grid leak) and capacitor act as an automatic bias circuit. During positive swings the grid draws current (since it is more positive than the cathode) and charges the capacitor. During the remainder of the cycle, the capacitor discharges through the resistor. The time constant of the resistor-capacitor combination is much longer than a cycle, however, and the capacitor accumulates a charge in the direction shown in the drawing. Because of this, if the amplitude of oscillation increases, the capacitor acquires a greater charge and applies additional negative bias, which reduces the gain of the system. By proper adjustment of the grid leak and grid capacitor, as well as by proper choice of the voltage ratio in the feedback transformer, an operating point can be chosen at which the amplifier is reasonably linear so that the oscillator output has a good waveform.


TUNE TO OBTAIN PROPER PHASE SHIFT AT DESIRED OPERATING FREQUENCY
Rp
B +
B−

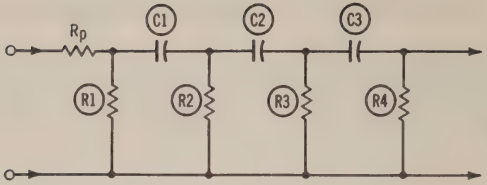
(A) Simple transformer circuit.

(B) Oscillator circuit.

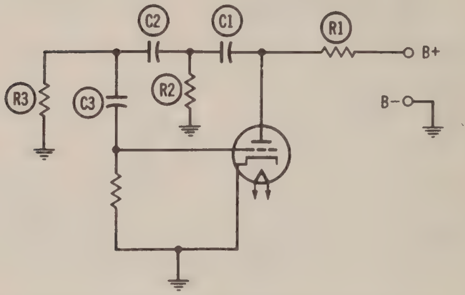
Fig. 5-14. Transformer-coupled oscillator circuit.

RC OSCILLATOR

Another simple load network that provides the required phase shift is composed of a number of resistor-capacitor sections (Fig. 5-15A). The circuit of the oscillator shown in Fig. 5-15B, is often used in the audio-frequency range.



(A) RC sections.

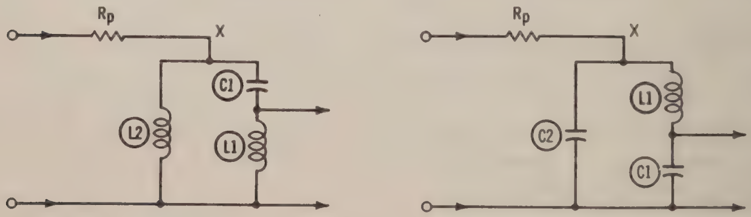


(B) Oscillator circuit.

Fig. 5-15. RC oscillator circuit.

THREE-REACTANCE OSCILLATORS

The three-reactance networks shown in Fig. 5-16 is the most frequently used load network, particularly at medium and high frequencies. In Fig. 5-16A it is shown with a single capacitor, and in Fig. 5-16B it is shown with two capacitors. The theory of these networks is simple. If in Fig. 5-16A the reactance of C is greater than the reactance of L1, the branch C1-L1, will look capacitive. The current through the branch then will lead the voltage at *X* by 90°. The voltage across L1, however, always leads the current through L1 by 90°,



(A) Single-capacitor circuit.

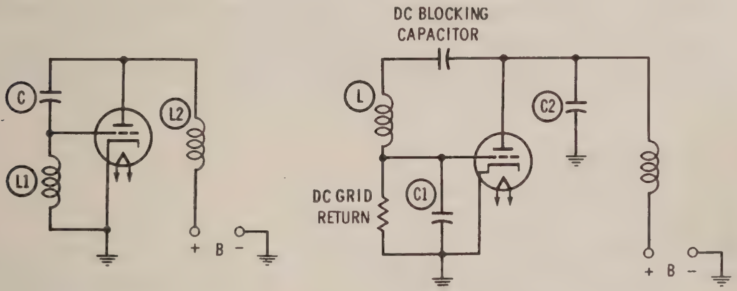
(B) Dual-capacitor circuit.

Fig. 5-16. Three-reactance networks.

and it is therefore 180° out of phase with the voltage at X. The output of the load network is therefore 180° out of phase with the voltage at X. If, then, L2 is chosen so that its inductive reactance is equal (but of opposite sign) to the capacitive reactance of branch C1-L1, the impedance at X will be resistive and the voltage at X will be in phase with the voltage at the input.

Similarly, in Fig. 5-16B, if the impedance of L1 is greater than that of C1, the branch will look inductive. The current in branch L1-C1 will therefore lag the voltage at X by 90° and, of course, the voltage across C1 lags the current by 90°. The result is a 180° phase shift between the voltage at X and the output voltage. If C2 is chosen properly, the impedance at X will be resistive at the oscillator frequency, and the network will provide the required 180° phase shift between input and output voltages.

The basic circuits of three-reactance oscillators are shown in Fig. 5-17: with one capacitor in Fig. 5-17A, and with two capacitors in Fig. 5-17B. The one-capacitor circuit uses series feed at both plate and grid, (the circuit element is in series with the DC supply to the tube element), but the two-capacitor circuit is forced to use shunt feed for both plate and grid (the circuit element is in parallel with the DC supply to the tube element). This requires a DC blocking capacitor in series with the plate-grid inductor, as well as an RF choke for the plate feed and a resistor for the grid feed.



(A) Single-capacitor circuit. (B) Dual-capacitor circuit.

Fig. 5-17. Three-reactance oscillator circuits.

Tuned-Plate, Tuned-Grid Oscillator

A tuned-plate, tuned-grid oscillator (Fig. 5-18) is a three-reactance oscillator with one capacitor. The two tanks, or frequency and to function as the inductors in a three-reactance tuned circuits, are adjusted to look inductive at the oscillator

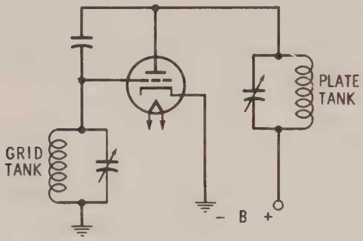


Fig. 5-18. Tuned-plate, tuned-grid oscillator circuit.

network. The third reactance is the capacitor between the plate and the grid. This capacitor is required when a pentode is used; but, when a triode is used, the plate grid capacitance of the tube is usually sufficient to complete the network, and the capacitor may be omitted.

Hartley Oscillator

The Hartley oscillator shown in Fig. 5-19 is another three-reactance oscillator with one capacitor. The two inductors, however, are wound as a single coil so that shunt feed must be used for the plate. This requires the RF choke and the DC

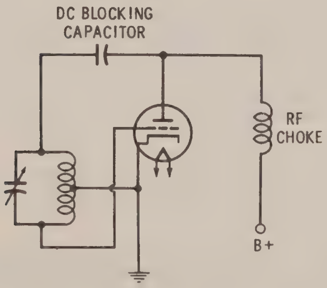


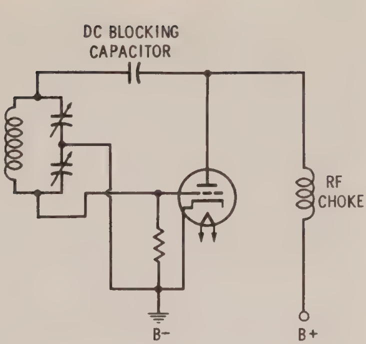
Fig. 5-19. Hartley oscillator circuit.

blocking capacitor shown in the drawing. The Hartley oscillator is simpler than the tuned-plate, tuned-grid oscillator because it has only one tuning adjustment.

Colpitts Oscillator

The Colpitts oscillator shown in Fig. 5-20 is a two-capacitor version of the three-reactance oscillator. In practice the two tuning capacitors are ganged so that only a single control is

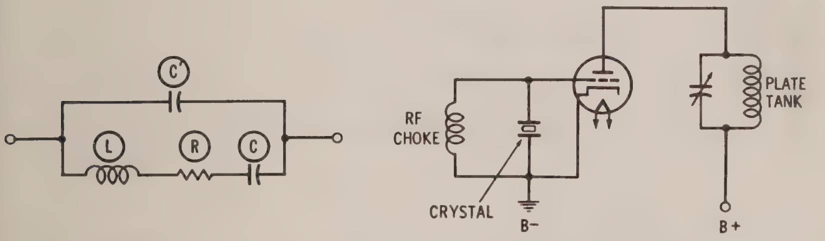
Fig. 5-20. Colpitts oscillator circuit.



needed. Since one side of each of the variable capacitors is grounded, the Colpitts oscillator is a little easier to construct from a mechanical standpoint than the Hartley oscillator.

Crystal Oscillator

A crystal oscillator is one that uses a quartz crystal as a frequency-controlling element. (This should not be confused with the crystal mixer and crystal rectifiers used at microwave frequencies.) The equivalent circuit of the crystal is shown in Fig. 5-21A, and the oscillator circuit is shown in Fig. 5-21B. This is a three-reactance oscillator of the one-capacitor type and is electrically equivalent to the tuned-plate, tuned-grid oscillator described previously and shown in Fig. 5-18. The required capacitance between the plate and the grid is supplied by the tube capacitance. The crystal that is shunted by the RF choke provides a DC grid connection. It looks inductive at the operating frequency; the plate tank is tuned to look inductive as well.



(A) Equivalent circuit.

(B) Crystal oscillator.

Fig. 5-21. Crystal oscillator circuit.

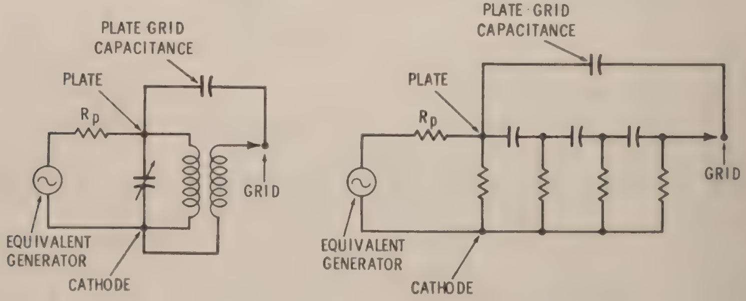
The precision with which the quartz crystal controls the oscillator frequency is indicated in Fig. 5-21A. At the resonant frequency of the crystal, the LRC (inductance-resistance-capacitance) branch of the equivalent circuit is resistive. Below the resonant frequency, however, this branch looks capacitive. In either case, the equivalent circuit of the crystal, the LRC branch in parallel with C', looks capacitive. Above the resonant frequency of the crystal, the LRC branch looks like an inductance. The crystal, however, can appear inductive only if the inductive current through the LRC branch is *greater* than the capacitive current through C'. Because of the high *Q* of the crystal, the inductive reactance of the LRC branch increases very rapidly with frequency (above the resonant frequency of the crystal) and there is only a small frequency range in which oscillation is possible. The *Q* of a crystal is something like 10,000. (This is much higher than the *Q* of an ordinary tuned circuit.)

TUBE CAPACITANCE AND LEAD INDUCTANCE

The feedback oscillator discussed in the previous paragraph appears to be a simple and uninvolved circuit. The situation is complicated, however, by the presence of capacitance between the elements of the vacuum tube and, to a lesser degree, by the inductance in the leads to the tube elements.

Transformer-Coupled and RC Oscillators

Neither the transformer-coupled nor the RC oscillator can be operated successfully, except at relatively low frequencies. In these oscillators, as in all feedback oscillators, signals are



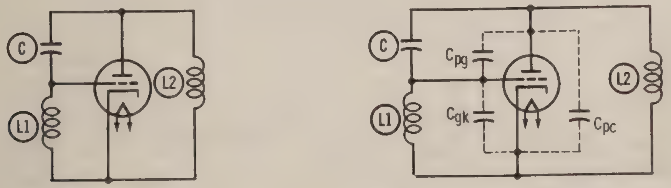
(A) Transformer-coupled. (B) RC coupled.

Fig. 5-22. Oscillator circuits showing plate-grid capacitance.

transferred from the plate to the grid through the load network. However, some additional signal is also transferred from the plate to the grid through the plate-grid capacitance of the tube; this path does not provide the required 180° phase shift in either transformer-coupled or RC oscillator circuits. Fig. 5-22A shows the circuit of a transformer-coupled oscillator making use of the equivalent generator concept with the plate-grid capacitance presented as though it were an external capacitor. Fig. 5-22B is the circuit of an RC oscillator. These circuits will oscillate only when the effect of the plate-grid capacitance is very small. Since the reactance of the plate-grid capacitance decreases when the frequency is raised, high-frequency operation of either of these circuits is difficult or impossible because the bulk of the grid signal is supplied through the plate-grid capacitance rather than through the phase-shifting load network. The resulting grid signal, of course, is not of the proper phase to sustain oscillations.

Three-Reactance Oscillators

Crystal oscillators cannot be used much above 10 mc, because of the extremely thin crystals required for high-frequency operation. (Crystal-controlled transmitters are operated at high frequencies, but these involve oscillators followed by harmonic generators.) Other types of three-reactance oscillators can, however, be operated at frequencies close to the microwave region.



(A) Basic circuit. (B) Showing tube capacitances.

Fig. 5-23. Three-reactance oscillator.

The three-reactance oscillator with one capacitor is attractive for high-frequency operation since the plate-grid capacitance becomes part of the load network and the other two tube capacitances may be tuned out by external inductances. Fig. 5-23A shows the basic circuit of the oscillator; Fig. 5-23B represents the circuit with the interelectrode capacitances added. Direct-current connections are omitted in each case.

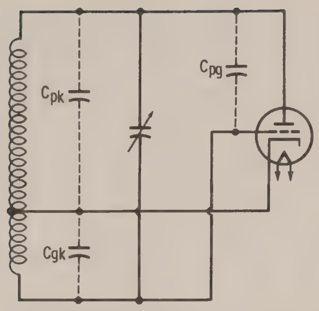
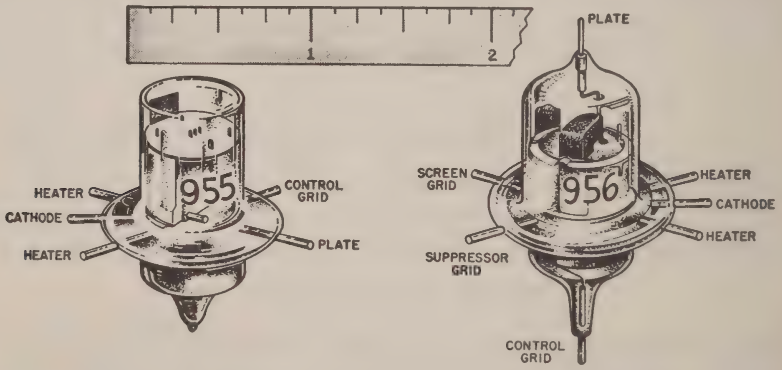


Fig. 5-24. Hartley oscillator circuit showing interelectrode capacitances.

The conventional Hartley oscillator, using tubes having standard bases (Fig. 5-24) with interelectrode capacitances added, operates well at moderately high frequencies. However, the possibility of operating such a Hartley at very-high frequencies is limited because the external inductors are in series with the inductances of the tube leads. It is therefore impossible to obtain *little* enough inductance for VHF operation using ordinary tubes.

This difficulty was solved in the middle 1930's by the development of the acorn tube. The early acorn tube series consisted of a triode, type 955, shown in Fig. 5-25A; and a sharp cutoff pentode, type 956, shown in Fig. 5-25B. The two pentodes are identical in external appearance, and use indirectly heated cathodes operating at 6.3 volts. Since the time of their development, other acorn tubes have been developed for battery operation. Interelectrode capacitances are minimized by small elements, and lead inductance is minimized


1
2
955
HEATER
CATHODE
HEATER
CONTROL GRID
PLATE
PLATE
956
SCREEN GRID
SUPPRESSOR GRID
HEATER
CATHODE
HEATER
CONTROL GRID

(A) Triode type.

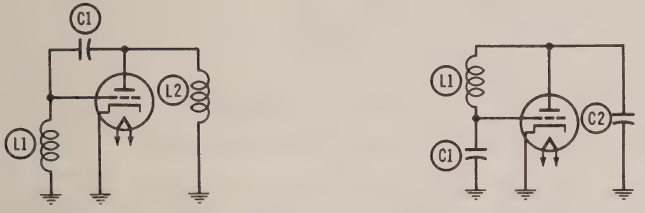
(B) Pentode type.

Fig. 5-25. Two types of acorn tubes.

by short straight leads. By virtue of this construction, the acorn triode can be used in a Hartley oscillator circuit at frequencies as high as 600 megacycles.

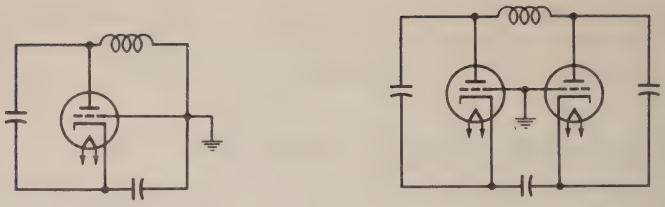
High-Power Three-Reactance Oscillators

Besides limiting the interelectrode capacitance of the tube, the miniature construction of the acorn tube unfortunately limits the power-handling capabilities as well. A large tube having greater allowable dissipation than the acorn and developing a correspondingly greater power output cannot be operated at such high frequencies. Fig. 5-26A shows the basic three-reactance oscillator with one capacitor. It is clear from the circuit (discussed earlier) that the required phase shift in branch C1-L1 can be obtained only if the impedance of the branch is capacitive—that is, if the impedance of C1 is greater than that of L1. Since the plate-grid capacitance is always present, this sets an upper limit on the frequency of the one-capacitor type of oscillator, unless the circuit is modified by the addition of a shunt inductor that can be used to raise the plate-grid impedance. A simpler solution is the use of the two-capacitor circuit shown in Fig. 5-26B. It is convenient to operate the grid instead of the cathode at ground (Fig. 5-27A). The reason for this is illustrated in the push-pull circuit in Fig. 5-27B. The midpoint of the plate-to-plate inductor is essentially at ground potential from the RF standpoint; thus, the inductor is connected from plate to ground. This is, of course, the same as a plate-grid connection if the grid is grounded. In practice, short lengths of transmission line may be used for the plate-to-plate inductor and the cathode-to-cathode capacitor, the lengths being chosen to obtain the proper impedance. Interelectrode capacitance in each tube supplies the plate-cathode capacitance. Fine tuning is accom-



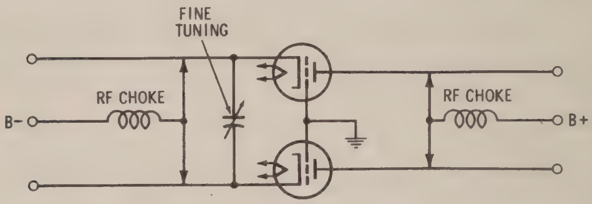
(A) Inductive output. (B) Capacitive output.

Fig. 5-26. Three-reactance oscillator circuits.



(A) Basic circuit.

(B) Push-pull with grounded grid.


FINE TUNING
RF CHOKE
B−
RF CHOKE
B+

(C) High-power oscillator.

Fig. 5-27. High-power oscillator circuit.

plished by means of a low-capacitance trimmer across the cathode transmsision line. An oscillator of this type, operated with a plate supply of 10,000 volts, is able to deliver up to

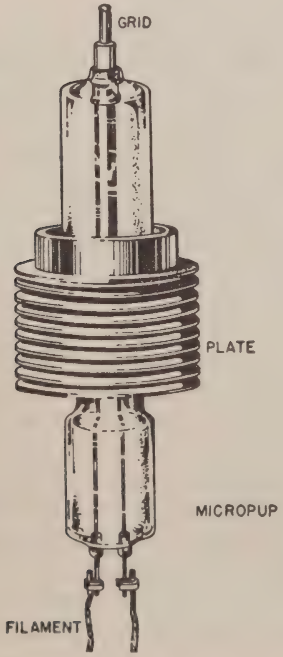

GRID
PLATE
MICROPUP
FILAMENT

Fig. 5-28. High-power, high-frequency triode.

50 kw in pulse service (during pulses) at frequencies of 200 mc. The tubes especially developed for this application are commonly called micropup tubes (Fig. 5-28). A number of other conventional tubes are able to operate at frequencies higher than this. High-power oscillators that are similar to the one just described can be operated easily at frequencies up to 600 mc, and acorn tubes can be operated above 1,000 mc.

Transit Time

This discussion has shown that interelectrode capacitance and lead inductance are likely to be troublesome when an attempt is made to operate a conventional oscillator at high frequencies. However, with proper circuit design and the use of special tubes (which are conventional vacuum tubes so far as operating principles are concerned) the basic three-reactance, feedback oscillator can be made to deliver useful power at frequencies above 1,000 mc. The limiting frequency for a given tube, above which operating is not feasible, is usually set by the *transit time*—the time it takes an electron to travel from the cathode to the plate.

Transit-Time Effect—Because of transit time, the grid of a vacuum tube operated in the conventional way draws a resistive current. Thus, the effect of transit time on the external circuit is identical with the effect of a resistor connected between the grid and the cathode of the tube. Referring to Fig. 5-29 assume that the plate-cathode voltage and the grid-cathode voltage are constant. Assume, further, that the grid is sufficiently negative so that no grid current flows. The plate current reaches a constant value shortly after the plate-cathode and grid-cathode voltages are applied, and, after this constant value has been reached, the number of electrons approaching the grid is the same as the number of electrons moving away from it.

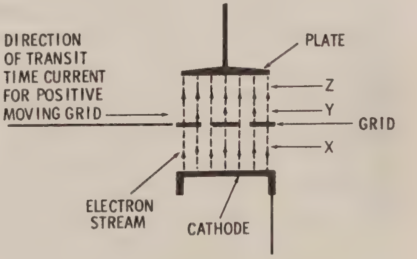


Fig. 5-29. Effect of electron transit time on vacuum-tube operation.

Now suppose that the grid is suddenly made more positive, but not so positive that grid current flows after the steady-state plate current has been established. Almost immediately the current flow will increase—more electrons will be moving in regions X and Y close to the grid. However, a short period of time will elapse before the current flow in region Z increases. This time delay results from the fact that the grid can establish increased current flow only in its immediate neighborhood. Therefore, the current flow at Z cannot increase until the arrival, in that region, of the electrons that leave the grid at the instant when the grid voltage is changed. After the new value of plate current has been established, the number of electrons approaching the grid is the same as the number receding from it. Before the new value of plate current has been established, while the current at Z still has its original value, more electrons are approaching the grid than are moving away from it. This is true because the current change in region Z, which is the last change to occur (the current changes earlier in regions X and Y because they are closer to the grid), increases the number of electrons moving away from the grid and makes this number equal to the number of electrons approaching the grid. *The excess of approaching electrons constitutes a grid current, whether or not the approaching electrons ever reach the grid.* These electrons repel other electrons at the grid and cause an actual current of electrons leaving the grid by way of the grid lead. An increase in the grid voltage is, therefore, accompanied by a momentary flow of current *to* the grid (or a flow of electrons from the grid) in exactly the same manner that an increase in the voltage across a resistor and a capacitor in series causes a momentary current flow. The effect of transit time is therefore the same as the effect of a series RC circuit connected between the grid and the cathode. When an alternating voltage is applied to the grid, the phase and magnitude of the transit-time current depend on the relation between the transit time and the duration of one cycle. At moderate frequencies the effect of transit-time current is to increase the apparent grid-cathode capacitance. At higher frequencies the resistive effect is the important one, and the equivalent input resistance of the grid decreases with increasing operating frequency, other things being equal.

Effect of Transit Time on Oscillator Performance—Suppose that the amplifier which is the heart of a feedback oscillator, delivers a power output, P. Some of this power is dissipated in the load network that feeds the grid, some of it supplies the required grid-driving power, and the remainder appears as useful output. With increasing frequency, the input resistance of the grid is decreased, because of transit time effects, and the power required to maintain a given voltage level at the grid is increased. At some frequency the required grid-driving power that is added to the power dissipated in the load network will be greater than P; consequently, sustained oscillations will not be possible. That is, a tube output of P watts will neither provide the grid-driving power nor the grid voltage level required to develop an output of P watts. This is simply another way of saying that the loop gain is less than 1.

Methods of Decreasing Transit Time—There are two methods of decreasing transit time and thus increasing the maximum frequency of oscillation. One method is to reduce the spacing between the tube elements, particularly between the grid and the plate, so that the electrons have a shorter distance to travel. Reducing the spacing without making any other changes would increase the interelectrode capacitances; therefore it is necessary to make a corresponding reduction in the size of the tube elements in order to hold the capacitances within reasonable limits. This method (reduction in spacing and in size) is the one followed in the design of acorn and lighthouse tubes. Unfortunately, reduction in the size of the tube elements effects a corresponding reduction in the power dissipation of the tube so that use of this method tends to produce tubes having only low-power outputs. The second method, used in high-power transmitting tubes, is to increase the plate-cathode voltage so that electrons move from cathode to plate more rapidly. This method is followed in micropup and similar tubes. Plate voltages of 10,000 volts are common, and even higher voltages are not unusual.

Lighthouse Tubes

Most of the conventional high-power oscillator tubes used at frequencies near the microwave region are essentially similar to the micropup tube. They are triodes that rely on close spacing of the elements *and* on high anode voltages for reduction of transit time. In the low-power field, however, the acorn tubes have been superseded, for operation at the highest frequencies, by disk-seal or *lighthouse* tubes.

In most vacuum tubes both the grid and the anode are in the form of open-ended cylinders that surround the cathode. Instead, the lighthouse tube uses a series of plane or flat ele-

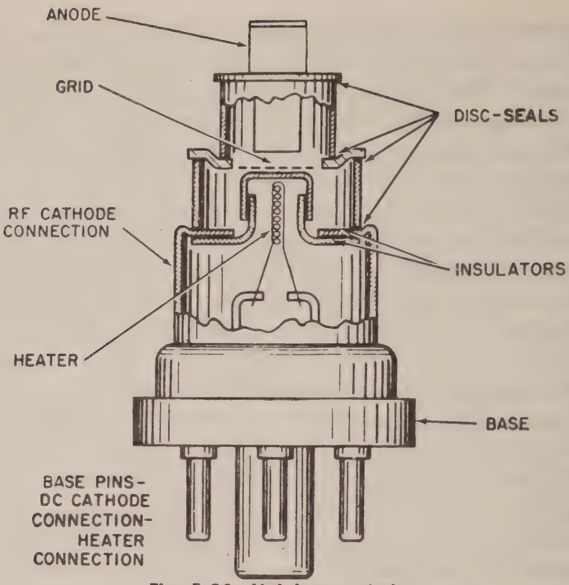


Fig. 5-30. Lighthouse triode.

ments. The resemblence of this tube to a lighthouse is slight, but nevertheless supplies its name.

A cross section showing the construction of this tube is shown in Fig. 5-30. The DC cathode connection is through the base pins, and the RF cathode connection is by way of the largest shell. The internal insulator between this shell and the cathode serves as the dielectric in a blocking capacitor, which keeps DC voltage from the shell. The use of disc terminals reduces the lead inductance of the lighthouse tube to extremely low values. In addition, the internal construction allows the grid conductor to extend beyond the cathode and anode so that it acts as an RF shield between them. By means of this design the capacitance between the cathode and the anode is reduced to a small fraction of a micromicrofarad.

The lighthouse tube is generally used in a grounded-grid circuit. The anode-grid impedance is made inductive, the cathode-grid impedance is made capacitive (and adjustable) and the anode-cathode capacitance of the tube itself forms the third reactance of the feedback network. The tuning elements are ordinarily rigid coaxial lines (Fig. 5-31). The outer conductor and the middle one form the cathode-grid line, and the middle and inner conductors form the anode-grid line—both lines are tuned by means of adjustable shorting plungers. The output is taken from a coupling loop or a probe (not shown) which extends into the space between the anode and

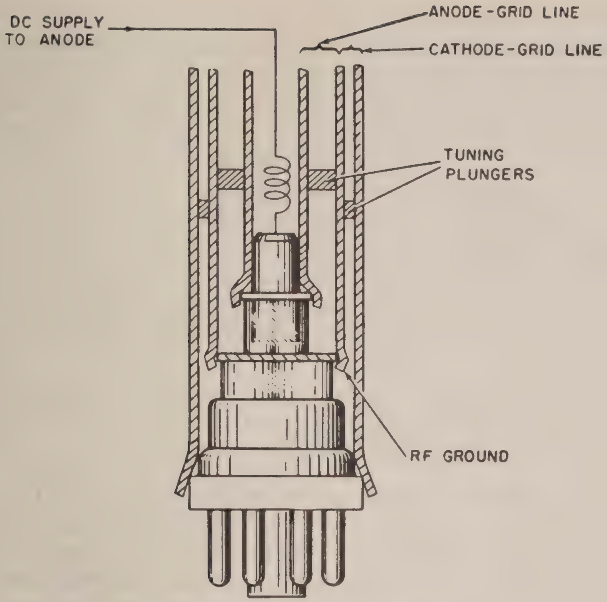


Fig. 5-31. Lighthouse oscillator.

the grid conductors. The lighthouse tube is frequently used as an oscillator at the lower microwave frequencies, although it can also be used as an RF amplifier in a receiver.

Barkhausen Oscillator

Another type of oscillator, called the Barkhausen oscillator, also uses conventional vacuum tubes. It is mentioned here because of its similarity to the klystron oscillators used at microwave frequencies.

Consider a conventional triode in which the anode is at the same voltage as the cathode and the grid is at some positive potential. Electrons emitted by the cathode will, of course, be attracted to the positive grid. The grid, however, is constructed of very fine wires with large spaces between them, and most of the electrons will pass through the spaces. After passing through the grid structure, the electrons continue to be attracted by the grid, since it is more positive than the anode, and the electrons reverse their direction and approach the grid again. The grid is therefore surrounded by a cloud of electrons, all of them oscillating through the plane of the grid (Fig. 5-32). As fast as electrons are captured by the grid and removed from the cloud, new electrons that have been emitted by the cathode take their places. As explained in the discussion on transit time, an electron approaching the grid is equivalent to a current *to* the grid, and an electron leaving the grid is equivalent to a current *from* the grid. There are oscillatory currents of all possible phases in the grid line. The frequency of these, which is twice the frequency of the electron oscillations, is determined by the electrode voltages and by the spacing between the electrodes. The frequency of the oscillatory currents is twice that of the electron oscillations, because the current direction is reversed four times during each electron oscillation cycle—once at each end of the electron path when the direction of the electron motion is reversed, and twice at the midpoint of the electron path when the electron passes through the grid structure.

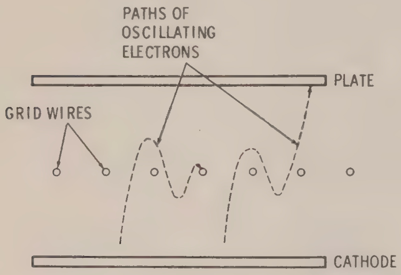


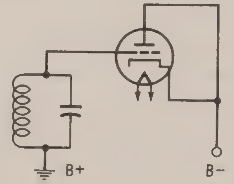
Fig. 5-32. Barkhausen oscillations.

Electron Sorting—Assume that an AC voltage having the same frequency as the oscillatory currents is superimposed on the positive DC grid voltage. These electrons that approach the grid during the positive half-cycle of the AC grid voltage and that leave the grid during the negative half-cycle, develop resistive currents and draw power from the generator that feeds the grid. That is, current associated with these electrons flows *to* the grid during positive half-cycles of the grid voltage and *away from* the grid during negative half-cycles. Electrons that oscillate 180° out of phase with these, however, deliver power *to* the grid, since the associated currents flow *from* the grid during positive half-cycles of the alternating grid voltage and *to* the grid during negative half-cycles. Therefore these electrons deliver power to the generator that excites the grid.

First, consider what happens to one of the electrons that delivers power to the grid. During the positive half-cycle of the grid voltage, this electron is moving away from the grid; during the negative half-cycles, it is moving toward the grid. Therefore the electron is attracted strongly when it is leaving the grid and only weakly when it approaches the grid. The strong attraction as it moves away prevents it from escaping from the neighborhood of the grid, and the weak attraction as it approaches the grid prevents it from being captured immediately. It continues to oscillate for a relatively long time. Now, consider the behavior of an electron that absorbs power from the generator that excites the grid. This electron is strongly attracted as it approaches the grid and only weakly as it leaves the grid. Consequently, it may be captured by the grid, or alternatively, it may escape the influence of the grid entirely and travel either to the anode or the cathode. The result is that the electrons that deliver power to the grid continue to oscillate longer than those that absorb power from the grid.

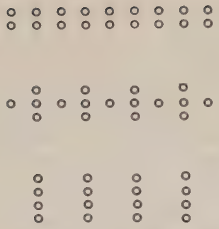
The circuit of a typical Barkhausen oscillator is shown in Fig. 5-33. Any momentary inequality in the oscillatory currents in the grid line causes the resonant circuit to ring. This, in turn, delivers an AC voltage to the grid, which shorts out the oscillating electrons according to the scheme described, removing those electrons that absorb power. The oscillating electron cloud continues to deliver energy to the resonant circuit, and stable oscillations result. The initial excitation of the resonant circuit results from tube noise—-a slight variation in the number of emitted electrons which is present in all tubes.

Fig. 5-33. Barkhausen oscillator circuit.



VELOCITY MODULATION

Velocity modulation, like almost everything else connected with microwaves, at first glance appears to be complicated, but actually it is fairly easy to understand. Briefly, if a beam of electrons passes a fixed point, the nature of the beam may be derived from the number of electrons that pass the point during a given interval of time, and the velocity with which the electrons travel as they pass the point. In a conventional vacuum tube operating in a conventional circuit, the electron beam is modulated by varying the number of electrons. In a velocity-modulated tube, however, the beam is modulated by varying the electron velocities.



(A) Unmodulated stream.

(B) 50% modulated stream.

(C) 100% modulated stream.

Fig. 5-34. Examples of density modulation.

To clarify the concept of velocity modulation, men are used in the following example (Fig. 5-34) as a substitute for electrons. Assume that a policeman passes prisoners through a gate on their way to the dining hall. As an example of an *unmodulated* beam, two men at a time pass through the gate at 1-second intervals, as in Fig. 5-34A, and all are marching at the same speed. At the gate, and at any other point between the gate and the dining hall, a steady stream of men will pass at the rate of two per second. As an example of *density-modulated* beam, in which the number of prisoners is varied, suppose that the policeman continues to pass prisoners through the gate at 1-second intervals, but, at this time, he passes one prisoner, then three, then one, then three, and so on. In radio terms, the stream of men now is density-modulated. An observer sees men passing alternately at rates of one per second and three per second, as in Fig. 5-34B. The modulation level is, of course, 50%. That is, the maximum number per second is 50% greater than the number per second in the unmodulated stream, and the minimum number per second is 50% less than that in the unmodulated stream. A modulation level of 100% is achieved if four prisoners pass through the gate at 2-second intervals, four at a time, as shown in Fig. 5-34C. The rate is then alternately zero and four per second.

As an example of a *velocity-modulated* beam, suppose the men pass through in pairs at 1-second intervals, as in the unmodulated stream, but that the first pair of men march more slowly than usual, the next pair slower, the next pair faster, and so on. At the gate, the stream appears to be unmodulated but, as the men march, the slow ones fall back and the faster ones move ahead in the stream. At some point, each pair of fast-marching prisoners will overtake the slow pair ahead and, at this point, they will pass an observer in groups of four, one group every 2 seconds (Fig. 5-35). At some further point, each faster pair will overtake the next slower pair ahead of it, and an observer will see the same thing. Therefore, if the velocity-modulated stream is observed at the proper point, and there are an infinite number of proper points, it will appear to be density-modulated.

Applegate Diagram

The behavior of a velocity-modulated beam is commonly shown on a graph called an *Applegate diagram* (Fig. 5-36). Fig. 5-37A shows the distance traveled by an electron as a function of time. The electron passes a given point (which is the source of the velocity-modulated beam) at a time that is taken as t = 0. It then travels with constant velocity so that its distance from the point increases uniformly with time. Fig. 5-37B shows similar plots for a number of electrons that leave the given point at intervals after the time t = 0. Suppose, now, that the beam is velocity-modulated. Some elec-

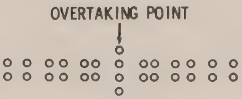

OVERTAKING POINT

Fig. 5-35. Example of velocity modulation.

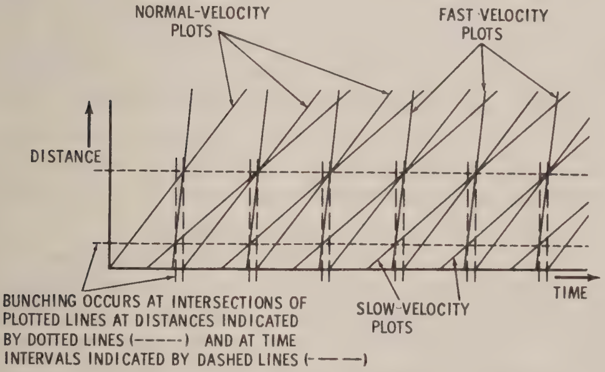
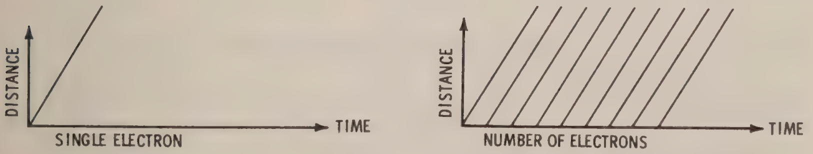

NORMAL-VELOCITY PLOTS
FAST VELOCITY PLOTS
DISTANCE
TIME
BUNCHING OCCURS AT INTERSECTIONS OF PLOTTED LINES AT DISTANCES INDICATED BY DOTTED LINES (-----) AND AT TIME INTERVALS INDICATED BY DASHED LINES (-——)
SLOW-VELOCITY PLOTS

Fig. 5-36. Applegate diagram.


DISTANCE
TIME
SINGLE ELECTRON
DISTANCE
TIME
NUMBER OF ELECTRONS

(A) Single electron.

(B) Multiple electrons.

Fig. 5-37. Electron time-distance plots.

trons will travel faster than others, and the plots for those electrons will be more nearly vertical than the plots for the slower ones as shown in Fig. 5-36. The points where bunching occurs—where the faster electrons overtake those traveling at normal velocity and where the slower electrons fall back to those traveling at normal velocity—are clearly shown in this diagram.

Some Velocity-Modulation Calculations

Consider a beam of electrons that is velocity-modulated at a frequency of f cycles per second. Let V be the speed of the standard-velocity electrons, let $V(1 + k)$ be the speed of the fastest electrons, and let $V(1 - k)$ be the speed of the slowest electrons. Finally, let T be the duration of a cycle, so that $T = 1/f$. Suppose that one of the fast electrons leaves the beam source at a given instant. The slow electron immediately ahead left the beam source 1 half-cycle earlier, and has already traveled a distance $V(1 - k)\ T/2$. The preceding slow electron left the beam source 3 half-cycles earlier, and has traveled three times as far, and so on. The fast electron moves at a speed of $V(1 + k)$, and the slow electron move at a speed of $V(1 - k)$. Therefore the fast electron overtakes all of the slow electrons, and its overtaking speed is $2kV$. The time required to overtake the first slow electron is $V(1 - k)\ T/2$ divided by $2kV$, or $(1 - k)\ T/4k$. The second slow electron is overtaken after a time interval three times as long as $(1 - k)\ 4k$; the third one, after a time interval five times as long; etc. The distance traveled by the fast electron is $V(1 + k)$ times the time interval so that the first slow electron is overtaken after the fast electron travels a distance $V(1 + k)\ (1 - k)\ T/4k$ or $V(1 - k^2)\ T/4k$. The second slow electron is overtaken after three times this distance, etc. Therefore, bunching takes place at distances from the beam source that are *odd* multiples of $V(1 - k^2)\ T/4k$. Since k is usually small, the quantity $(1 - k^2)$ is nearly equal to 1, and the bunching distances may be taken as odd multiples of $VT/4k$.

KLYSTRON OSCILLATORS

A simple klystron consists of four parts: a beam source, a velocity-modulating unit called a buncher, and a drift tube through which the velocity-modulated beam travels, and a unit called a catcher that takes energy from the velocity-modulated beam.

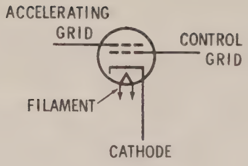


Fig. 5-38. Klystron beam source.

The klystron beam source shown in Fig. 5-38 consists of a cathode and its heater, a control grid, and an accelerating grid. The control grid, as is usual, controls the number of electrons in the beam. It does *not*, however, have any signal superimposed on the bias voltage. The accelerating grid speeds up electrons that are passed by the control grid. Velocity V is adjusted by varying the voltage on the accelerating grid.

The *buncher* (Fig. 5-39) is a resonant cavity, usually of modified cylindrical form, that has two bunching grids through which the electron beam passes. When microwave energy is fed to the buncher cavity (the input to the cavity is not shown in Fig. 5-39) oscillations are set up, and an alternating voltage appears between the grids. When the second grid is more positive than the first, electrons passing through the buncher drift tube are speeded up. Conversely, when the second grid is more negative than the first, electrons passing through the buncher are slowed down. Therefore, the electron beam is velocity-modulated as it passes through the buncher (provided that the buncher cavity is fed with microwave energy). It is necessary, of course, to tune the cavity, and for this purpose tuning paddles and plugs are provided.

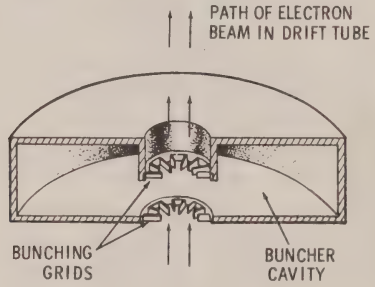


Fig. 5-39. Klystron buncher.

The cavity is made re-entrant in order to reduce the spacing between the two bunching grids. Thus, if an electron remains in the space between the grids for a whole cycle, it will be speeded up during part of the cycle and slowed down during the remainder. The spacing is reduced and the electron velocity is made large so that the time required for an electron to pass between the buncher grids is only a small fraction of a cycle.

The *drift tube* (Fig. 5-40) is simply an evacuated tube through which the electron beam moves, and the *catcher* is a cavity with two grids like the buncher. In practice, the buncher and the catcher are almost identical.

In Fig. 5-40 all parts of a complete klystron are shown separately, including a collector electrode on which the electrons land after passing through the catcher grids. In practice, the cathode is made negative, and the remainder of the tube, with the exception of the control grid, is operated at ground potential. The second catcher grid frequently serves as a collector, and the first buncher grid often serves as an accelerating electrode. Operation of the klystron is simple. An electron beam is passed through the buncher and is velocity-modulated by feeding energy to the buncher cavity. As the beam travels down the drift tube, the electrons are bunched; when the tube is adjusted properly, bunching occurs at the

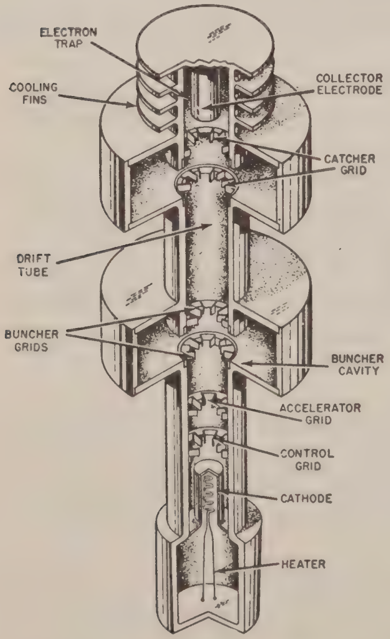

ELECTRON TRAP
COOLING FINS
COLLECTOR ELECTRODE
CATCHER GRID
DRIFT TUBE
BUNCHER GRIDS
BUNCHER CAVITY
ACCELERATOR GRID
CONTROL GRID
CATHODE
HEATER

Fig. 5-40. Cutaway view of a klystron tube.

catcher grids. Electrons then pass through the catcher grid in bunches at the microwave frequency. It has been pointed out that an electron current is equivalent to any other current, and an alternating current between the grids of the catcher will excite the catcher cavity. Therefore the tube acts as an amplifier, taking energy in at low levels at the buncher cavity and delivering an output from the catcher cavity. If some energy is fed back by means of a short length of coaxial cable from the catcher cavity to the buncher cavity, the klystron will oscillate when it is adjusted correctly.

It is necessary to tune both resonant cavities of the klystron before it will operate either as an amplifier or as an oscillator. Besides this, however, it is necessary to insure that bunching takes place at the catcher grids. This is controlled by adjusting the negative voltage on the cathode, thus maintaining a fixed voltage difference between the cathode and control grid so that the accelerator brings the electrons to the proper velocity, V.

As already stated, bunching occurs approximately at odd multiples of distance VT/4k, where the source of the modulated beam is the second buncher grid. Therefore the distance between the buncher and the catcher must be an odd multiple of VT/4k for the tube to amplify or oscillate. Otherwise, it will neither amplify nor oscillate since the electron beam delivers little or no energy to the catcher cavity. The klystron is said to be operating in the first mode (or in the fundamental mode) when the distance is approximately VT/4k; in the second mode when the distance is 3VT/4k; in the third mode when the distance is 5VT/4k; and so on. Due to collisions between the electrons bunching is less precise in the higher modes; fast electrons slow down on colliding, and slow electrons gain speed. Consequently, a klystron is normally operated in either the first or the second mode. It is possible to amplitude-modulate the klystron by varying the voltage on the control grid and to frequency-modulate it by changing the accelerating voltage. In practice, however klystrons of this sort seldom are used, since the reflex klystron is smaller and more efficient.

The klystron can function as an amplifier, as mentioned above, but it seldom is used because of a fundamental difficulty. For successful amplification, the distance between the buncher and the catcher must be an odd multiple (more or less) of distance VT/4k. Unfortunately, k changes with the input level and with constant spacing. As a result the klystron may be a good amplifier at one level and a poor one for signals at the same frequency at another level. This characteristic, which makes the klystron of almost no value as an amplifier, gives it stability as an oscillator.

Consider the mechanism of oscillation: some roughness in the electron beam excites the buncher cavity directly, or excites the catcher cavity and, through the coupling cable, the buncher cavity. In either case, the buncher cavity rings at its resonant frequency and velocity-modulates the electron stream. At a low signal level, *k* is small and the bunching distance is well beyond the catcher grids (since VT/4k is larger than it should be). If *k* is large enough—if the shock excitation of the cavity has been sufficiently severe—the klystron will amplify, and the signal will be fed back from the catcher to the buncher. As the signal level is increased, *k* becomes larger, the bunching distance becomes more correct, and the amplification of the klystron increases. At some point, the bunching of the electrons takes place at the catcher grids and the klystron operates at maximum gain. Further increases in *k* move the bunching distance away from the catcher grids again (still toward the buncher grids, by making VT/4k smaller) and reduce the amplification. Just as in a conventional feedback amplifier, the level of oscillation increases (and the loop gain decreases) until a stable operating level is reached with a loop gain of 1.

As an oscillator, the klystron operates well, seeking and maintaining a stable operating level. The klystron can be amplitude-modulated by changing the control-grid voltage (which alters the number of electrons in the velocity-modulated beam). It also can be frequency-modulated by changing the acceleration voltage. Considerable interaction occurs between the two types of modulation, however, and neither type is completely successful because there is no easy method of varying the phase shift through the feedback loop (the coaxial cable from catcher to buncher). In practice, the simple klystron has been displaced almost completely by the reflex klystron, which is smaller and simpler.

Before discussing the reflex klystron, it is desirable to look briefly at the simple klystron from another slightly more complicated point of view. Consider the resonant system, which consists of the two cavities (buncher cavity and catcher cavity) coupled by a coaxial cable. The discussion of the Barkhausen oscillator has shown that acceleration of an electron absorbs energy from the resonant system, and that deceleration feeds energy to the system. The resonant system then absorbs energy from all electrons that pass the catcher grids when the second catcher grid is less negative than the first. Because the velocity-modulated electron beam is bunched by the time it reaches the catcher grids, almost all of the electrons arrive at times when they can feed energy to the resonant system. If the phasing between the cavities is correct for the spacing and for the operating potentials, only a few electrons arrive at times when they absorb energy. Velocity modulating in the klystron is, therefore, equivalent to electron-sorting in the Barkhausen oscillator. On this basis, fast electrons, during acceleration, absorb energy from the buncher and feed it back to the catcher. Slow electrons, however, feed energy to both cavities. Successful operation of the klystron thus depends on the proper relation between cavity tuning, cavity spacing, phase shift through the coaxial cable, electron velocity, and modulation level. It will be seen next that the reflex klystron is simpler—this is its major merit.

Reflex Klystron

A reflex klystron may be thought of as a simple klystron that has been folded back on itself so that one cavity acts as both buncher and catcher (Fig. 5-41). It also may be considered as a type of Barkhausen oscillator with a cavity resonator for a tuned circuit. It consists of a cathode and heater, a resonator cavity with two grids, and a reflector.

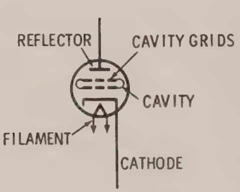


Fig. 5-41. Reflex-klystron tube schematic.

In the type 723A/B, which is the usual X-band reflex klystron, the cavity is tuned by physical distortion. The cavity, as well as the body of the tube, is operated at ground potential. The cathode is negative so that the first cavity grid serves as both accelerator and control grid. The second cavity grid, on the side away from the cathode, modulates the electron beam. The reflector is operated at a voltage slightly more negative than the cathode, and it serves to turn electrons back to the cavity.

Operation of the reflex klystron depends on a proper relation between the cathode voltage, the reflector voltage, and the cavity tuning. Consider for the moment what happens to an electron that passes through the cavity grids at an instant when the second grid is at the same potential as the first and is going negative. The electron is repelled by the reflector and finally is brought to rest. After this, its velocity is reversed and it moves back to the cavity grids, passing through and rejoining the cloud of electrons about the cathode. An electron that passed through the cavity grids slightly earlier (when the second grid is positive) approaches the reflector at a higher velocity and takes longer to stop. The result is that it arrives at the cavity grids, following reflection, after a longer trip. Similarly, an electron that passes through the cavity grids later than the first electron considered (when the second grid is negative) returns after a shorter trip. The result is a bunching process that is brought about by velocity modulation of the beam. Suppose, now, that a standard electron, which is not accelerated by the cavity grids, returns in ¾ cycle. An electron that passes the grids earlier, and is accelerated, returns in more than ¾ cycle. An electron that passes the grids later, and is slowed down, returns in less than ¾ cycle. As a result, if the system is tuned correctly, energy is delivered to the cavity, and the tube oscillates.

As in the case of the simple klystron, it is possible for a reflex klystron to operate in a number of different modes. In the first mode, an unaccelerated electron returns to the cavity grids in ¾ cycle. In the second mode, it returns after 1¾ cycles. In the third mode it takes 2¾ cycles to return, and so on. The frequency is controlled almost entirely by the reflector voltage, if the cathode voltage is held fixed; tuning the cavity puts the system in and out of oscillation. As with the simple klystron, the operating level is fixed by the fact that gain is reduced when the level of oscillation rises above that level for which bunching occurs at the cavity grids. The system seeks and maintains the level of oscillation at which the energy input to the cavity just balances the sum of the losses and the energy output, to the electron beam *and* to the load.

Modulation Techniques—When the reflector is made more negative in relation to the cavity grids, the slowing force on the electrons is increased. Therefore an electron passing through the cavity grids on its way from the cathode will return sooner and the frequency of oscillation will be raised. Of course, there is a limit to the amount of frequency change that can be obtained by changing the reflector voltage. If the cavity is initially tuned to resonance, it will be slightly out of tune when the frequency is changed by variation of the reflector voltage. This will cause the output of the oscillator to drop. By the time the frequency has been shifted by about 20 or 30 mc, the power will have dropped to about one-half of the maximum value obtained when the cavity is tuned. The frequency change is approximately 1 mc for a 1-volt shift in the reflector voltage. This variation in frequency is, of course, distinct from the change that can be made by changing the reflector voltage and retuning the cavity. In general, by changing the reflector voltage and returning the cavity a reflex klystron can be tuned over a range of about +5% from the center of its operating band. By changing the reflector voltage it can be frequency-modulated over a range of about ±.5% of its operating frequency. Note, however, that the frequency modulation is accompanied by some amplitude modulation when the frequency shift is accomplished by variation in the reflector voltage.

When the cathode is made more negative in relation to the cavity grids, electrons are accelerated more before reaching the grids. Therefore, they take longer to return to the cavity grids, and the frequency is lowered. At the same time, however, the number of electrons in the beam is increased so that the power output rises with decreasing frequency and falls with increasing frequency. The increase in the number of electrons in the beam when the cathode is made more negative is similar to the effect of an increase in the plate current of an ordinary tube when the plate is made more positive.

By changing both cathode and reflector voltages (although not by equal amounts) it is possible to achieve either amplitude or frequency modulation of a reflex klystron. If, for example, the cathode is made more negative, the power output increases and the frequency decreases. A slight increase in the reflector voltage will return the frequency to the initial value, or will prevent a frequency shift if the voltage changes are made simultaneously without affecting the change in power level. Similarly, if the reflector voltage is altered to frequency-modulate the klystron, a slight change in the cathode voltage will compensate for cavity detuning and hold the output level constant.

Pulsed Operation—The reflex klystron will not operate if either the cathode or the reflector is at the same voltage as the cavity grids. It may, therefore, be pulse-modulated by applying negative voltage pulses to either the cathode or the reflector. It is usual, in order to avoid frequency variation during pulses, to supply voltage pulses to both the cathode and reflector. Amplitude and frequency modulation of the pulsed output can be obtained by suitable modulation of the DC pulses (negative) that are applied to these electrodes.

Application of Reflex Klystrons—Reflex klystrons are used whenever microwave signals at low-power levels are required. Microwave signal generators, for example, use reflex klystrons with electrical tuning by variation in the reflector voltage, and with mechanical tuning by tuning of the cavity. Microwave receivers, which are almost always superheterodynes, use reflex klystrons as local oscillators. It is usual to take advantage of the possibility of electrical tuning by adding AFC (automatic frequency control) systems. These are described in Chapter 6, which deals with microwave receivers.

Another interesting application, which makes use of electrical tuning, is the spectrum analyzer. This is a device that sweeps a band of frequencies and displays the signal level (at each frequency) on an oscilloscope. The spectrum analyzer is nothing more than a narrow-band receiver, the output of which is fed to the vertical deflection plates of an oscilloscope.

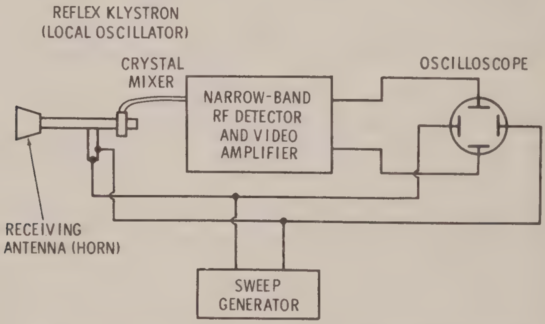


Fig. 5-42. Functional diagram of a spectrum analyzer.

An external sweep generator feeds a signal to the horizontal deflection plates of the oscilloscope and, at the same time, applies a sawtooth voltage wave, superimposed on the fixed DC voltage, to the reflector of the klystron which acts as a local oscillator for the receiver (Fig. 5-42). The result is a picture on the oscilloscope screen in which frequency is shown on a horizontal scale (the horizontal deflecting voltage is also the tuning voltage for the receiver), and signal level at each frequency is shown on a vertical scale.

CAVITY MAGNETRONS

Before discussing cavity magnetrons, it is desirable to examine the behavior of some simpler systems which consist of several resonant cavities, such as the buncher and catcher cavities of a non-reflex klystron, and a single velocity-modulated electron beam.

The two-cavity system (Fig. 5-43) consists of two resonant cavities, both tuned to the same frequency, and an electron beam which passes through both sets of cavity grids. Fig. 5-43 does not show the source of the electron beam, nor the evacuated envelope, since the system is of theoretical interest only. Suppose that the first cavity is excited at its resonant frequency by an external microwave power source connected to the coaxial input line. The electron beam will, of course, be velocity modulated. If the system is adjusted properly, with correct electron velocity and modulation level and correct spacing between the cavities, the beam will be bunched at the grids of the second cavity. Under these conditions, the second cavity will resonate at the operating frequency, and a sizable signal will appear there. Also, the phase of the oscillation in the second cavity must be such that the bunched electrons are slowed down, for this is the only way that energy can be abstracted from the beam. Oscillation at any other phase either will be unaffected by the electron beam, if the bunched electrons pass through the cavity grids when both are at the same potential, or will deliver energy from the cavity to the beam, if the bunched electrons pass through the cavity grids at a time when they are accelerated. The electron beam,

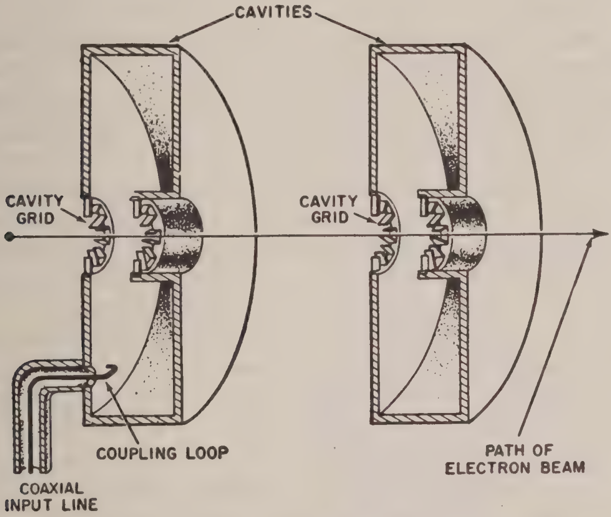


Fig. 5-43. Two-cavity system with electron-beam coupling.

]

therefore, is able to carry energy from the first cavity to the second.

Consider now what happens if the second cavity is moved toward the first. The electron beam then is underbunched at the grids of the second cavity; that is, bunching is not yet complete. Since there is partial bunching, the flow of electrons is not smooth, and the second cavity is able to absorb energy from the beam. The phase of the oscillation in the second cavity automatically adjusts itself, by absorbing energy, so that the greatest possible number of electrons is slowed. The remaining electrons, which pass the cavity grids when the second grid is more positive than the first, are accelerated. The second cavity, therefore, slows down more than half of the electrons that pass through its grid, and accelerates less than half of them. Thus the cavity absorbs energy from the beam, reducing the average velocity of the electrons in the process (since it slows down more electrons than it speeds up) and also adds to the velocity modulation of the beam.

If a cavity system has more than two cavities, it is possible for each cavity, except the first one, to absorb energy from the beam and, at the same time, improve the velocity modulation of the beam for the next cavity. A small portion of the energy comes from the external source that feeds the first cavity. The greater part of the energy, however, is supplied by the electron beam source; each cavity reduces the average velocity of the electrons in the beam. Note that the system is extremely sensitive to adjustment, since the energy absorbed by each cavity depends on the relation between bunching distance and spacing between cavities.

Suppose, now, that a number of resonant cavities, all tuned to the same frequency, are arranged in a ring (Fig. 5-44). Assume, further, that a beam of electrons is made to travel in a closed loop that passes through the grids associated with each cavity. For the moment, the question of how the electron beam is made to take this path may be disregarded. Any

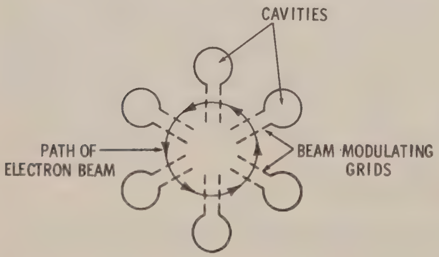
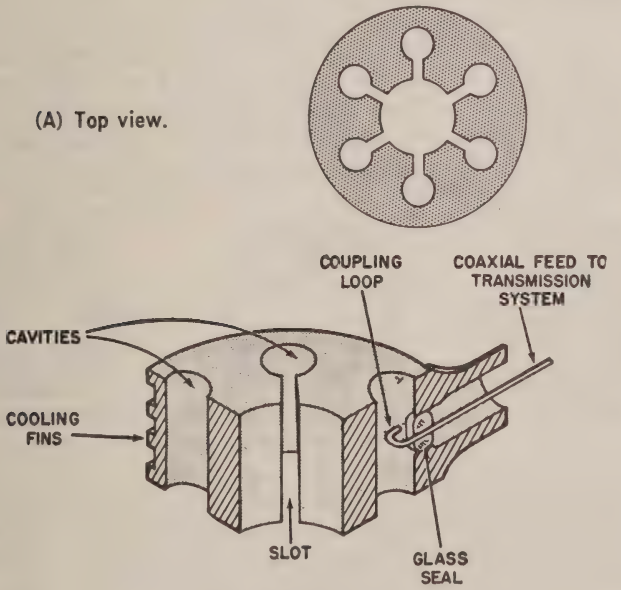


Fig. 5-44. Ring system of resonant cavities.

roughness in the electron stream will cause one or more of the cavities to ring. This will modulate the beam, and energy will be delivered to all the cavities. The microwave energy delivered to the cavities comes, as before, from the original velocity of the electron beam, and transfer of energy to the cavities results in a reduction of the average electron velocity. If matters can be arranged so that a continuous supply of fast electrons is available, while slow electrons are removed from the system, energy will be abstracted from the electron stream and will appear in all of the cavities. (The energy in the cavities is, of course, RF energy in the form of electric and magnetic fields). The cavity magnetron operates in precisely this way.

A cavity magnetron consists of a cylindrical anode in which a series of cavities are cut, a cathode, and a magnet. The magnet controls the paths of the electrons so that they travel in a more or less circular orbit. Fig. 5-45 shows one of the more common magnetron anodes—a hole-and-slot type. All of the types of magnetron anode are basically similar. Fig. 4-45A is a view along the axis of the cylinder that shows only the main outline of the anode. The reason for the name



(A) Top view.

(B) Perspective view.

Fig. 5-45. Hole-and-slot anode.

is obvious at once, since each of the cavities consists of a cylindrical hole joined to the central space by a small slot. Although six cavities are shown, magnetrons with other numbers of cavities are not uncommon. Fig. 4-45B is a sectional view of the anode in perspective showing the cooling fins on the outer surface of the anode and also showing the coupling loop and coaxial line, which takes energy from one of the cavities to the microwave transmission system, either waveguide or rigid coaxial line. There are no cavity grids. Instead, the electron beam is modulated by the field, which spreads out from the lips of the slot, as shown in Fig. 5-46. Other types of anodes are shown in Fig. 5-47. Fig. 5-47A shows a slot anode, Fig. 5-47B, a vane anode, and Fig. 5-47C, a rising-sun anode. Choice of one anode type over another is based largely on the problems associated with construction of the magnetron, since all types behave in the same way.

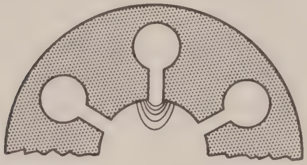
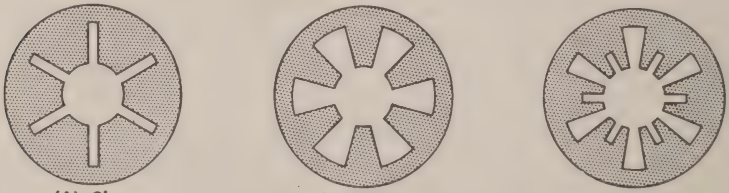


Fig. 5-46. Magnetic field in the neighborhood of the slot.


(A) Slot.
(B) Vane.
(C) Rising-sun.
Fig. 5-47. Types of anodes.

The cathode of the cavity magnetron is placed along the axis of the anode cylinder. It is usually indirectly heated, and is connected electrically to one of the heater leads so that only two leads are required for heater and cathode (Fig. 5-48). Most magnetrons are used in pulse service, and the cathode

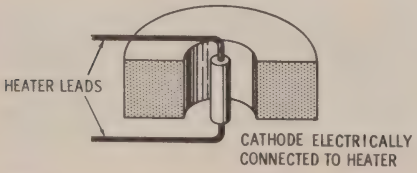
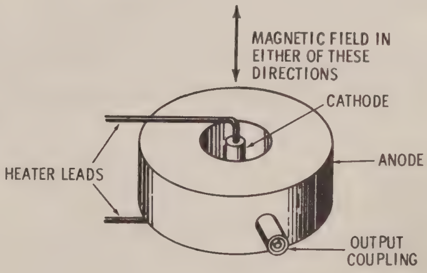


Fig. 5-48. Magnetron cathode.

emission during pulses is in the neighborhood of 50 amperes, so a high emission oxide is almost always used. It will be shown that many of the emitted electrons ultimately return to the cathode, and the energy that they give up contributes to the cathode heating in the same way that the anode of a conventional vacuum tube is heated by the energy it absorbs from the electrons it receives. Because of this, the heater current is made high during warm-up of the equipment and is reduced during operation. In some high-power magnetrons, the heater is even turned off during operation, since the electrons captured by the cathode supply enough energy to maintain it at normal operating temperature. The magnetron *never* must be operated without allowing sufficient time for the cathode to come to operating temperature. If this precaution is not observed, the magnetron may be damaged.

The magnet of the cavity magnetron is usually separate from the cathode and the anode, and the term *magnetron* normally is applied to the cathode-anode structure and the enclosing envelope. Fig. 5-49 shows a simplified view of the magnetron structure showing the direction of the magnetron field, parallel to the cathode.

Fig. 5-49. Direction of magnetic field in the magnetron.



A magnetron magnet is usually supplied with a keeper—a cylinder of soft iron which should be placed between the pole pieces of the magnet whenever the magnetron is removed. *It is necessary to resist the temptation to play with the magnet* by bringing screw drivers and pieces of metal near, since this weakens it and will affect the performance of any magnetron that subsequently is used with it. Do not wear a wrist watch when working with a magnetron magnet, for even if it is of the nonmagnetic type, it will probably be damaged. The poles of the magnet normally are marked North and South, and although the magnet may be placed around the magnetron in either position, one position may be preferable to the other. If such is the case, the magnetron will be marked, and it should be inspected for a notation of this sort before putting it in service.

Theory of Operation

For its operation, the magnetron depends on the fact that the path of an electron is deflected when the electron moves at right angles to a magnetic field. As a matter of interest, the same principle is used in every cathode-ray tube that has magnetic deflection (Fig. 5-50).

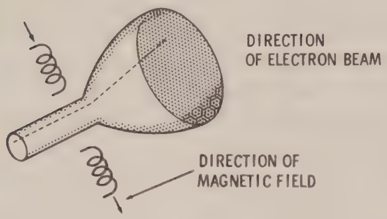


Fig. 5-50. Magnetic deflection in a cathode-ray tube.

Aside from the cavities in the wall of the anode structure, the magnetron may be considered as a cylindrical *diode with* a uniform magnetic field parallel to its axis. In the absence of the magnetic field, the magnetron would behave like any other cylindrical diode; the electrons would travel radially from cathode to anode, as in Fig. 5-51. However, a warning is inserted here: *Do not operate a magnetron without the magnet in place* as this would ruin the magnetron.

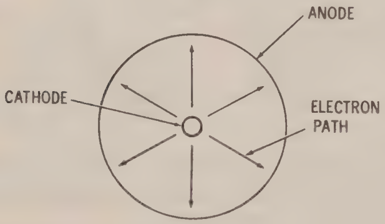


Fig. 5-51. Electron paths without magnetic field.

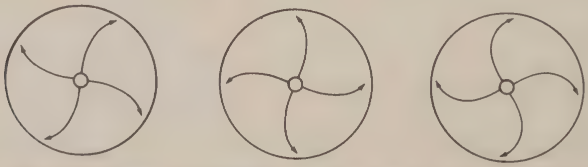


Fig. 5-52. Effect of moderate magnetic field on electron paths.

In Fig. 5-52, A shows the effect of a moderate magnetic field on the electron paths. This field, however, is much too low for normal operation. Note that the paths all are curved in the same direction by the uniform magnetic field, which is perpendicular to the plane of the page. If this field is replaced by one in the opposite direction, the electron paths will resemble those shown in B. A stronger magnetic field, or a lower voltage between cathode and anode, will lead to the paths shown in C, which are curved more sharply. When the correct relation exists between the magnetic field strength and the cathode-anode voltage, the electron paths are as those shown in A of Fig. 5-53. This is a rather idealized picture, since it neglects all interaction between the electrons and also neglects the effect of the cavities. In practice, collisions between the electrons disrupt the simple path that runs from cathode to anode and back again. The actual path spirals outward so that the electrons ultimately are captured by the anode, as shown in B of Fig. 5-53.

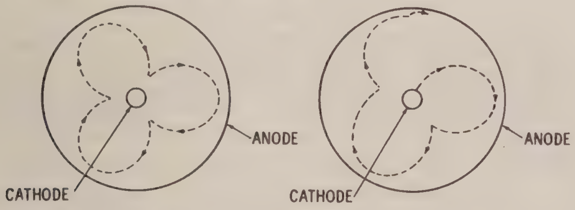


Fig. 5-53. Electron paths with proper magnetic field.

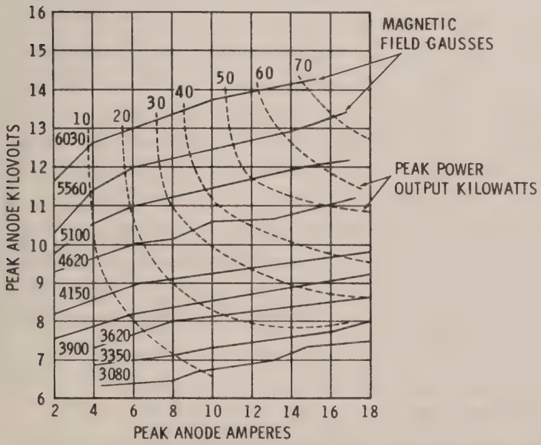


Fig. 5-54. Variation of power output and anode current with change in magnetic field strength and anode voltage.

Although paths of the proper sort can be obtained with any anode voltage if the appropriate magnetic field is used, or with any magnetic field if the correct anode voltage is used, the possibility of oscillation (transfer of energy from the electrons to the cavities) depends on the relations between electron velocity. This relation is a function of the anode voltage, spacing between the cavities, and resonant frequency of the cavities. The magnetron will, therefore, oscillate over only a relatively narrow range of magnetic field strength and anode voltages. Fig. 5-54 shows the relation between anode voltage, anode current, magnetic field strength, and power output for a typical magnetron operating near 9,000 mc, in the lower part of the X-band. A graph of this sort is known as a performance chart.

Operation of the magnetron is not involved. Because of the magnetic field, electrons pass by the resonant cavities and transfer energy to them. As electrons slow up, they are removed from the system by the anode. The picture is complicated, although only slightly, by the presence of the end plates. Fig. 5-55 shows a simplified sketch of a magnetron, showing the end plates, which form a portion of the wall of the evacuated space. In practice, the body of the magnetron is operated at ground potential and the cathode is anywhere from 10,000 to 50,000 volts negative. In pulsed operation, negative DC pulses are supplied to the cathode by a high-power pulse generator called a modulator. A voltage difference, therefore, exists between the cathode and the end plates, and some of the electrons wander off to the end plates without delivering much energy to the cavities. Changing the position of the end plates alters the electron paths and also, to some extent, changes the resonant frequencies of the cavities. Advantage is taken of this fact in tunable magnetrons, whose operating frequency can be varied by about 10% through distortion of the end plates.

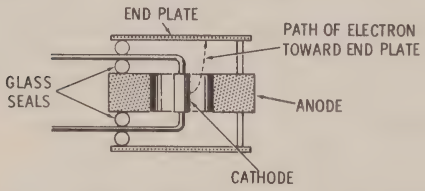


Fig. 5-55. Effect of end plate on electron paths.

Modes of Operation

It is possible for a magnetron to operate in a number of different modes. In the case of a six-cavity magnetron, for example, the phase difference between adjacent cavities may be 60°, in which case the sum of the six phase differences is 1 cycle. If the phase difference is 120°, the sum is 2 cycles; or if 180°, the sum is 3 cycles. Similarly, in an eight-cavity magnetron, the phase difference between adjacent cavities may be 45°, 90°, 135°, or 180°. The desired mode is the one with phase differences of 180°, and other modes frequently are suppressed by means of *strapping*. The strapping for a six-cavity hole-and-slot anode is shown in Fig. 5-56. Fig. 5-56A shows an end view of the straps, and Fig. 5-56B shows a perspective view. The straps simply connect corresponding points on adjacent cavities. Since the straps have an electrical length of ½ wave, a phase difference of 180° is assured. Only one lip of each cavity is shown strapped in Fig. 5-56A, but the other lips are strapped at the other end of the anode (not shown). Other strapping arrangements are possible, of course, and all the straps may be at one end of the anode, or there may be a complete set of straps at each end. A magnetron of this sort is called a strapped magnetron.

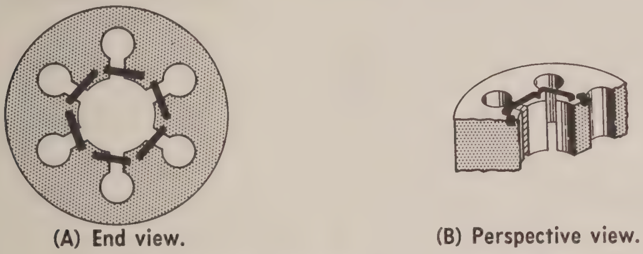

(A) End view. (B) Perspective view.

Fig. 5-56. Physical construction of a strapped anode.

In view of the complex nature of the magnetron oscillator, it is not surprising to find that both frequency and power output vary when the load impedance is changed. Unfortunately, the variation in frequency for a given change in load is greatest when the load is adjusted for maximum power output from the magnetron (the input being held constant). It is therefore customary to adjust the load for a minimum VSWR at the magnetron rather than for maximum magnetron output. A graph of power output and frequency as a function of VSWR usually is called a Rieke diagram, shown in Fig. 5-57 for a 2J50 magnetron. Since the power output and frequency for a given VSWR may take on any of the values on the corresponding VSWR circle, the only way of achieving reproducible operating conditions is to tune the transmission system for minimum VSWR at the magnetron.

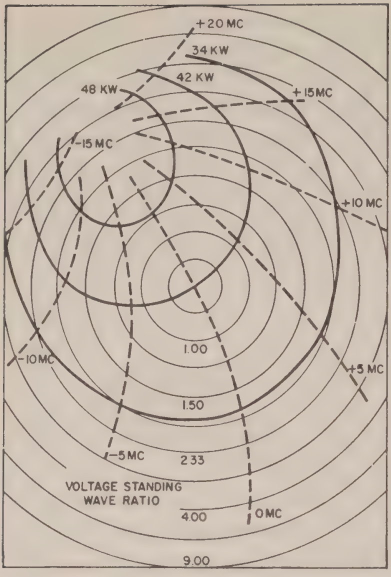


Fig. 5-57. Rieke diagram.

Magnetrons are available in sizes up to 1,000 kilowatts output for pulse service with a duty cycle of about 1/1,000 (in operation only 1/1,000 of the time). Higher duty cycles, and even continuous operation, are available in lower-powered magnetrons. In general, as the operating frequency is increased, the structure of a magnetron must be made smaller. The corresponding reduction in heat dissipation ability lowers the available power output. Most magnetrons operate with efficiencies of approximately 35%, and, consequently, any reduction in dissipation capability has a marked effect on the power input that can be used, and on the power output. It is, therefore, important to make sure that blowers or other cooling systems are operating properly. Magnetrons can be modulated by variation of the anode voltage, but this results in both amplitude and frequency modulation, and is generally unsatisfactory. For this reason, various forms of pulse width and pulse position modulation are used in high-power microwave systems which use magnetrons. Considering, now, the problem of stability rather than of modulation, most magnetrons are stable to better than one-half of 1% in frequency with variation in load. There are various methods of achieving greater stability through the use of external high-*Q* cavities, but these seldom are used. Microwave receivers are inherently broad-band in any event, and slight frequency variation in a microwave transmitter can be tolerated. If necessary, the receiver can be provided with an automatic frequency control system (Chapter 6), so that it follows the transmitter through most variations in frequency.

Chapter 6

Microwave Receivers

Almost all radar receivers are of the superheterodyne type (Fig. 6-1). Incoming signals are passed through an RF section, which may or may not contain an RF amplifier, to a frequency converter. Here, the signals are converted by heterodyning with the output of a local oscillator, to a lower frequency that is called the intermediate frequency. The IF output of the frequency converter is amplified and then passed to a suitable detector, which is followed by appropriate video amplifiers. Since we are concerned with only microwave techniques, rather than with radar techniques in general, many

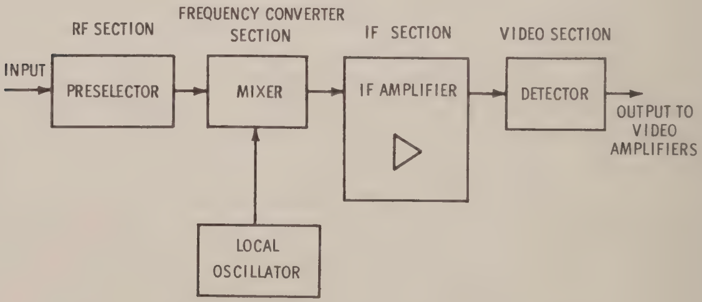

RF SECTION
FREQUENCY CONVERTER SECTION
IF SECTION
VIDEO SECTION
INPUT
PRESELECTOR
MIXER
IF AMPLIFIER
DETECTOR
OUTPUT TO VIDEO AMPLIFIERS
LOCAL OSCILLATOR

Fig. 6-1. Block diagram of a superheterodyne receiver.

aspects of radar receivers are touched on only lightly in the following discussions.

FREQUENCY CONVERSION

The heart of a superheterodyne receiver is the frequency converter, which receives signals at the operating radio frequency and delivers an output at the intermediate frequency. The frequency converter consists of a nonlinear element, or circuit, called a mixer, and a source of unmodulated signals, at a frequency close to that of the incoming signals, called a local oscillator. At VHF and lower frequencies, both the mixer and local oscillator are sometimes combined in a circuit which uses only a single vacuum tube. Such a circuit is called a converter.

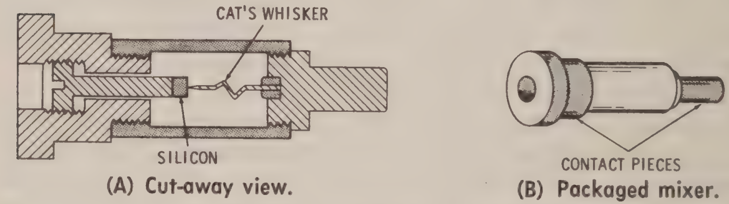
MICROWAVE FREQUENCY MIXERS

The mixer may be any nonlinear circuit which is capable of operation at the frequency in use. (A nonlinear circuit is one in which the output is not proportional to the input.) When a single signal of amplitude A and of frequency f is passed through a nonlinear circuit, the output not only contains input frequency f, but also harmonics of the input with frequencies 2f, 3f, 4f, etc., as well as some DC. When two signals of amplitudes A_1 and A_2 and frequencies f_1 and f_2 are fed to a nonlinear circuit, the output contains DC and harmonics of *both* of the input signals at frequencies $2f_1$, $3f_1$, $4f_1$, etc; *and* $2f_2$, $3f_2$, $4f_2$. In addition, the output contains signals at frequencies $(f_1 + f_2)$, $(f_1 - f_2)$, $(2f_1 + f_2)$, $(2f_1 - f_2)$, $(f_1 + 2f_2)$, $(f_1 - 2f_2)$, etc. The fact that some of these differences lead to negative values is not important since a signal of frequency $(f_1 - f_2)$ and a signal of frequency $(f_2 - f_1)$ are essentially the same. The important component of the output, from the standpoint of frequency conversion, is the signal having a frequency $(f_1 - f_2)$. This signal has an amplitude proportional to (A_1A_2), if the mixer is designed properly.

In an S-band receiver, for example, one of the signals may be at a frequency of 3,000 mc (the frequency of the incoming signal) with amplitude P, which is proportional to the amplitude of the incoming signal. The other input to the mixer is the output of the local oscillator, which has a frequency of 2,970 mc and constant amplitude K. The desired component of the output has an amplitude proportional to (KP) and a frequency of (3,000 – 2,970) mc. This is, of course, a signal at a frequency of 30 mc (the intermediate frequency). Since *K* is constant, the signal has an amplitude proportional to that of the incoming signal. In other words, the incoming signal is reproduced (although usually at a higher level) at a frequency 2,970 mc lower than its original frequency. Some amplitude-modulation detectors operate in the same way, using the constant-amplitude carrier and the variable-amplitude sidebands. For this reason, a mixer or converter sometimes is called a first detector, whereas the actual detector is called the second detector.

Crystal Mixers

The usual mixer used at microwave frequencies is a crystal mixer. This consists of a thin wire, called a cat's whisker, which makes contact with a small piece of silicon or germanium. The arrangement is shown in Fig. 6-2A; the packaged mixer is shown in Fig. 6-2B. The nature of the metal, either silicon or germanium, is such that relatively few electrons are available for current carrying. Such a metal is called a semiconductor.



(A) Cut-away view. (B) Packaged mixer.

Fig. 6-2. Construction of a crystal mixer.

Although it is easy to inject electrons through the cat's whisker, it is difficult to take them out in any sizable quantity because of the small contact area. The result is that such a crystal passes current easily from the metal to the cat's whisker, and with difficulty in the opposite direction. Current flow is in the opposite direction from electron flow; current in a vacuum tube flows from plate to cathode, whereas electrons travel from cathode to plate. The current-voltage relationship for a crystal mixer is shown in Fig. 6-3. For applied voltage in the backward direction, the current is proportional to the voltage, and it is very small. For a large applied voltage in the forward direction, the current is proportional to the square of the applied voltage. The curved portion of the characteristic where the voltage-current relation is nonlinear

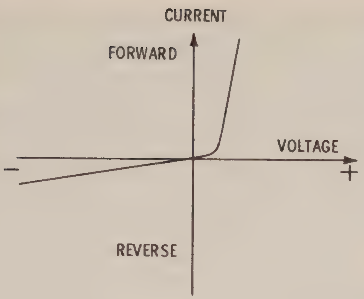
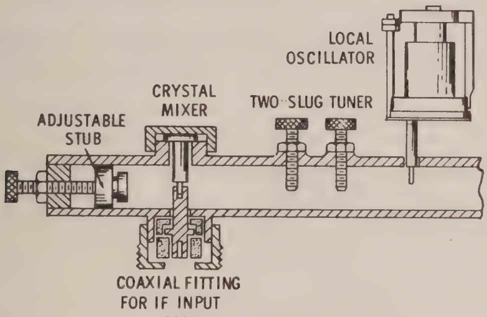
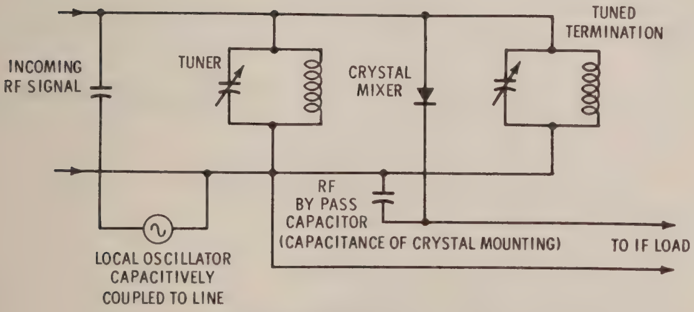


Fig. 6-3. Voltage-current relationship of a crystal mixer.

is used for mixing. The general construction of a waveguide frequency converter (local oscillator and mixer) is shown in Fig. 6-4A; the equivalent circuit is shown in Fig. 6-4B. The method of mounting the crystal mixer is shown clearly in the figure.


LOCAL
OSCILLATOR
CRYSTAL
MIXER
TWO-SLUG TUNER
ADJUSTABLE
STUB
COAXIAL FITTING
FOR IF INPUT

(A) Construction.


TUNED
TERMINATION
INCOMING
RF SIGNAL
TUNER
CRYSTAL
MIXER
RF
BY PASS
CAPACITOR
(CAPACITANCE OF CRYSTAL MOUNTING)
TO IF LOAD
LOCAL OSCILLATOR
CAPACITIVELY
COUPLED TO LINE

(B) Equivalent circuit.

Fig. 6-4. Waveguide frequency converter.

Since a crystal mixer is damaged easily by overloading, it is usually applied in an envelope of thin metal—frequently lead foil. The envelope serves the double purpose of protecting the mixer from moisture and shielding it from strong electric fields which might burn it out. When installing a crystal mixer, *touch the crystal mount with one finger just before inserting the crystal.* This discharges any static electricity that may have been generated, possibly by friction between shoe soles and floor, and guarantees that there will not be a short pulse of high voltage across the crystal.

Reference to Fig. 6-3 shows that the other voltage-current relation for a crystal mixer is the same as that of a diode (the crystal mixer often is called a crystal rectifier or crystal diode). The same sort of crystal is used as a microwave detector in connection with slotted-line sections and directional couplers. Any other diode may be used as a mixer, although pentagrid mixers are used at lower frequency because they amplify the signal in addition to shifting its frequency. The difficulty with most diodes is that the anode-cathode capacitance completely bypasses the diode at microwave frequencies. Aside from crystal mixers, the only mixer of any importance for operation at microwave frequencies is the lighthouse diode

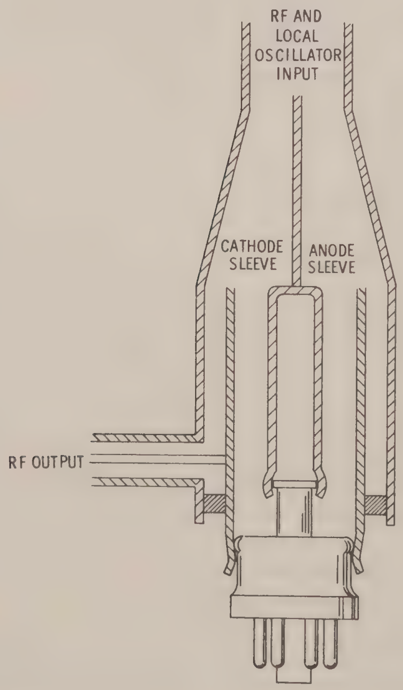


Fig. 6-5. Lighthouse diode.

type 559 shown in Fig. 6-5. The two signals, RF input and local oscillator input, are fed through the diode to the IF load. The section of transmission line consisting of the anode sleeve and the cathode sleeve is a tuning impedance which resonates the diode capacitance at the RF frequency. The section of line consisting of the cathode sleeve and the outer conductor acts as an RF bypass capacitor in parallel with the IF load. Mixers of this sort, however, seldom are used.

LOCAL OSCILLATORS

The usual local oscillator is a reflex klystron. Less frequently, and only on the S-band, a simple klystron is used. Occasionally, also on the S-band, a lighthouse local oscillator is used together with a lighthouse diode mixer.

FREQUENCY-CONVERTER ASSEMBLIES

A complete frequency-converter assembly, consisting of a TR switch, local oscillator, crystal mixer, and appropriate tuners, is frequently made as a single unit. A shutter seals off the crystal. When the equipment is turned on, the shutter is withdrawn by a solenoid after the TR switch is energized. At all times, therefore, the crystal mixer is protected by the shutter or a TR switch against damage by excessively strong signals.

RADAR BEACONS

In many applications, particularly in airborne operation, a single radar is used for both search and as a navigational

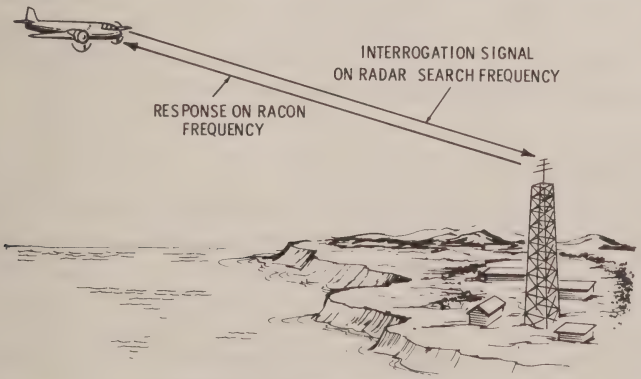


Fig. 6-6. Beacon operation.

instrument. Navigation information is obtained by interrogating radar beacons or *racons*. A racon is nothing more than a transmitter and a receiver connected in such a manner that the transmitter sends out a distinctive signal, or a series of signals, whenever the receiver picks up an interrogating signal (Fig. 6-6). The racon receiver *listens* on the frequency normally used by the search radar. The racon transmitter, however, *replies* on another frequency so that its signal may be easily distinguished from the normal radar echoes. Many search radar receivers are required to operate on either of two frequencies, one for search operations, the other for racon navigation.

SEARCH-AND-BEACON FREQUENCY CONVERTER ASSEMBLY

Frequency-converter assemblies are sometimes designed for operation on either of two frequencies. They differ from the simpler converter assembly just described by having two local oscillators. One is tuned for normal search operation and the other for reception on the racon band. Tuning from one frequency to the other is, then, a simple matter of turning one klystron off and the other on. For simplicity in tuning, a two-frequency assembly of this sort is usually provided with two crystal mixers and two TR switches, as well as two sets of tuning devices.

PRESELECTION

The RF amplifier, if used, is called a preselector in a superheterodyne receiver. At frequencies below the microwave region, it is customary for high-performance receivers to employ one or two stages of preselection in order to improve image rejection and receiver sensitivity, and to eliminate local oscillator radiation. At microwave frequencies, however, preselectors are used only in the S-band, and then only rarely.

Image Rejection

Consider a superheterodyne receiver designed to receive signals at frequency f_1, which uses an intermediate frequency f_2. If the local oscillator frequency is to be lower than that of the incoming signals, the oscillator will be tuned to a frequency $(f_1 - f_2)$, as in Fig. 6-7A. The difference between the frequency of the incoming signals and that of the local oscillator is, then, $f_1 - (f_1 - f_2)$, or f_2. The mixer will, however,

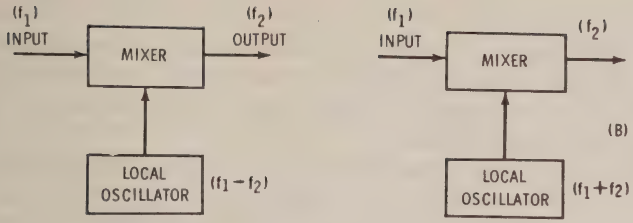


Fig. 6-7. Frequency converter.

deliver an output at intermediate frequency f_2, if there is an incoming signal at the *image frequency* $(f_1 - f_2)$, for the difference, $(f_1 - f_2) - (f_1 - 2f_2)$, is also equal to the intermediate frequency. The local oscillator is also operated at a frequency higher than that of the incoming signal, in which case, the local oscillator is tuned to a frequency $(f_1 + f_2)$, as in B. of Fig. 6-7. Here, the image frequency is $(f_1 + 2f_2)$. There is, then, an image frequency which is separated from the desired operating frequency by twice the intermediate frequency. Since the mixer will deliver an input at the desired frequency *or* with an input at the image frequency, all portions of the receiver after the mixer will treat signals at the image frequency in precisely the same way that they treat signals at the desired operating frequency. The only way of rejecting signals at the image frequency is by means of tuned circuits *ahead* of the mixer. The mixer input is normally tuned to the operating frequency, f_1; this offers some image rejection. The insertion of additional tuned circuits associated with one or two stages of amplification, at the RF frequency (f_1) offers greatly increased rejection of signals at the image frequency.

Receiver Sensitivity

In a superheterodyne receiver the output level for a given signal input is set by the overall receiver gain—by the amplification in the preselector, mixer, intermediate-frequency amplifier, and the audio or video amplifier. The selectivity, or ability to reject signals at frequencies just outside the desired bandpass, is determined by the number of tuned circuits in the intermediate-frequency amplifier. Image rejection is determined by the number of tuned circuits in the preselector. The receiver sensitivity, however, depends on the equivalent noise input at the first stage of the receiver. Every electrical circuit is troubled by noise. The noise in a circuit like a crystal mixer results from random motion of the electrons in the conductor. The noise in a vacuum tube results from roughness in the electron stream, as well as from random electron motion in the conductors of the grid circuit. On this account, every receiver has some output even when no signal is received from the antenna.

Suppose that, for a given receiver, this output averages 10 volts when the receiver is connected to a 100-ohm load. This is, of course, equal to 1 watt. If, now, an input signal of 5 microvolts raises the output voltage to 14 volts, which indicates a power output of about 2 watts, the signal-power output and the noise-power output are equal. The receiver noise, therefore, may be considered as the equivalent of a 5-microvolt signal, since it develops the same power at the receiver output. Although there are ways of detecting signals that are well below the level of the background noise, it is clear that the minimum usable signal is related to the equivalent noise input, since any reduction in the noise level will allow detection of weaker signals.

A preselector improves the sensitivity of a receiver (in general) because it has a lower equivalent noise level than a mixer, although this is not always true. Assume, for example, that a given receiver has an equivalent noise input of 5 microvolts, and that a preselector with a gain of 10 and an equivalent noise input of 2 microvolts is inserted ahead of the receiver. The noise at the mixer with the preselector, will be the original noise—equivalent to the 5-microvolt signal plus the preselector noise which, after amplification, is equivalent to a 20-microvolt signal. Then at the mixer, the level of the amplified preselector noise is 4 times as great as that of the mixer noise, and the power in the amplified preselector noise is 16 times as great as the power in the mixer noise. The mixer noise is, therefore, relatively unimportant and the equivalent noise input of the receiver is that of the preselector.

Local-Oscillator Radiation

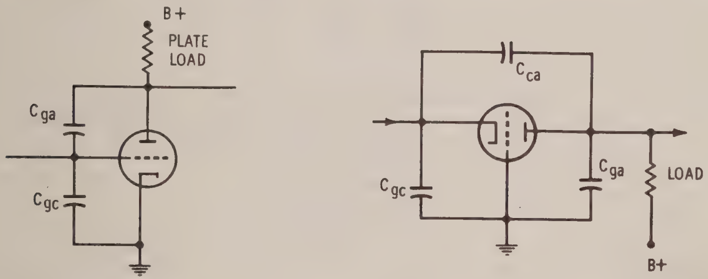
Whenever a mixer is connected directly to the transmission line from an antenna, it is possible for some of the signal from the local oscillator to travel along the same transmission line to the antenna. If this happens, the radiated signal may cause interference with other equipment. The insertion of an RF amplifier in the line between the antenna and the mixer eliminates this possibility, since the amplifier will only carry signals from the antenna but not to it.

Microwave Preselectors

Preselectors are not used at microwave frequencies in the X- and K-bands, since there is no practical method of amplifying signals at these frequencies. It is possible to use a simple klystron (not a reflex klystron) as an RF amplifier at S-band frequencies, but this is seldom done because the equivalent noise input of the klystron is higher than that of the crystal mixer. The only effective preselector is a lighthouse-tube amplifier and even this is seldom used, because it offers little advantage.

Grounded-Grid Amplifiers

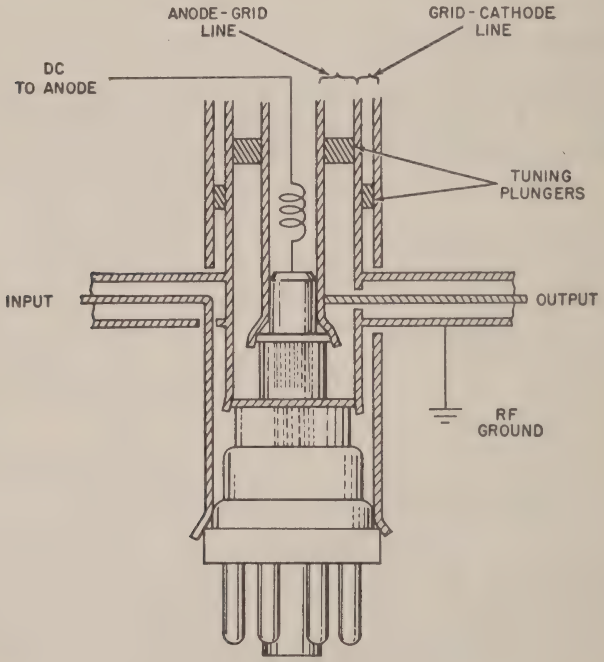
When a lighthouse tube is used an an RF amplifier at S-band frequencies, it is used in a *grounded-grid* circuit. The grid is grounded, the input signal is developed between cathode and ground, and the output signal appears between plate and ground. Consider the amplifier shown in Fig. 6-8A. This is a conventional circuit, except for the fact that grid-cathode capacitance C_{gc} and grid-anode capacitance C_{ga} are shown on the drawing. Suppose that the grid voltage is raised slightly. It is, of course, necessary to charge capacitor C_{gc} by delivering current from the driving source. If the anode potential remains fixed, capacitor C_{ga} must be charged from the same source, since the plate load is a high impedance for signal frequencies. The effective input capacitance of the tube therefore is the sum $(C_{ga} + C_{ga})$. That is, the circuit behaves exactly as though a capacitor of value $(C_{gc} + C_{ga})$ were connected from grid to ground. In fact, when the grid is made more positive, the anode moves in the negative direction *A* times as far (where *A* is the amplification of the stage). The voltage change across the grid-anode capacitance therefore is (1+A)



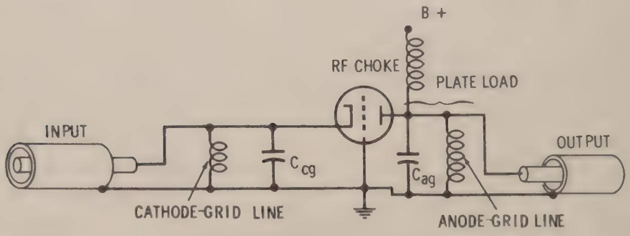
(A) Conventional. (B) Grounded-grid.

Fig. 6-8. Conventional and grounded-grid amplifiers.

times as great as the voltage change at the grid, and it requires (1+A) times as much charging current as it would if the plate voltage did not change. In short, it behaves like a capacitance *A* times as big as C_{ga} connected between grid and ground. The equivalent input capacitance of the circuit is, then ($C_{gc} + C_{ga} + AC_{ga}$). Suppose, now, that the grid is grounded, as shown in Fig. 6-8B. When the cathode is made more positive, it is still necessary to charge capacitor C_{gc}. The grid-anode capacitance, however, now is charged entirely through the plate load circuit. The cathode anode capacitance


ANODE-GRID LINE
GRID-CATHODE LINE
DC TO ANODE
TUNING PLUNGERS
INPUT
OUTPUT
RF GROUND

(A) Physical construction.


B +
RF CHOKE
PLATE LOAD
INPUT
OUTPUT
C_{cg}
C_{ag}
CATHODE-GRID LINE
ANODE-GRID LINE

(B) Equivalent circuit.

Fig. 6-9. Grounded-grid lighthouse amplifier.

must be charged, but this is much smaller than the grid-anode capacitance. In addition, if the cathode is made more positive, the plate also moves in the positive direction. With a gain of 1, both plate and cathode move through the same voltage change, and no charging current is required for C_{ca}.

A negative charging current is required with greater gain so that the current to charge C_{gc} is obtained in part, from C_{ca}. The resulting equivalent input capacitance is, then, only $(C_{gc} + C_{ca} - AC_{ca})$, and may be reduced nearly to zero.

A typical grounded-grid lighthouse amplifier is shown in Fig. 6-9A, with the equivalent circuit in Fig. 6-9B. The amount of noise in a receiver depends largely on the noise originated by nonuniformity in the electron flow in the first tube. If RF amplification is used, this tube is the first RF tube. Otherwise, it is the first IF tube in a crystal mixer. Instead of giving the equivalent input noise, it is sometimes convenient to compare the equivalent noise power input of the receiver with the noise power contributed by the antenna, which results from random motion of the electrons in space near the antenna. A noise figure of 1 would indicate that the receiver contributed no noise of its own. In practice, noise figures of 35 to 100 are usual. Except at the very lowest microwave frequencies, the noise figure for a receiver with an RF amplifier is likely to be higher than that for the same receiver with a crystal mixer. For this reason, preselectors seldom are used, even at the low microwave frequencies where the required RF amplifiers can be constructed. As stated above, preselectors are not used on the X- and K-bands because no RF amplifiers are available.

INTERMEDIATE-FREQUENCY AMPLIFICATION

The intermediate-frequency amplifier of a radar receiver is much like any other IF amplifier. The only specific requirement is that of bandwidth. For conventional, or pulse-type radar, it is customary to have a bandwidth of 3/d mc in the intermediate-frequency amplifier, where *d* is the pulse width in microseconds. Most search radars use pulse widths of approximately 1 microsecond, and adequate bandwidth can be obtained with an intermediate frequency of 30 mc. Tracking radars, in general, use shorter pulses and usually operate at an intermediate frequency of 60 mc. These two frequencies have become more or less standard for VHF and microwave operation; other intermediate frequencies seldom are encountered. Radar beacon receivers, which must pick up all interrogating pulses whether or not the interrogating transmitters are operating at the proper frequency, normally use an intermediate frequency of 60 mc in order to achieve greater bandwidth.

Bandwidths and Noise

As the bandwidth of a receiver is increased, the noise output also increases. The equivalent noise input, therefore, is proportional to the bandwidth. In a search radar, the basic problem is the detection of targets through the background of noise. The bandwidth is consequently reduced as far as possible in order to minimize the equivalent noise input. There is, however, a limit below which the bandwidth cannot be reduced without impairing receiver performance. With a band width of 1/d, signal and noise pulses are similar in appearance, and it is difficult to decide whether a small receiver output is the result of noise or of a weak signal. The customary bandwidth of 3/d passes more noise power than does a bandwidth of 1/d, but it also makes echoes more distinguishable from noise. The basic problem with tracking radars is the precise determination of range and, in some cases, echo amplitude. Since tracking radars operate at relatively short ranges and therefore with relatively strong signals the bandwidth frequently is increased to 5/d so that the echo signals may have sharper fronts. This, of course, permits the target range to be determined with greater precision.

Coupling From Mixer to IF Amplifier

The usual method of coupling to the intermediate-frequency amplifier is by means of coaxial cable from the mixer. The coaxial line is terminated in its characteristic impedance by a link inductively coupled to the tuned circuit at the grid of the first IF amplifier tube (Fig. 6-10). Other coupling methods

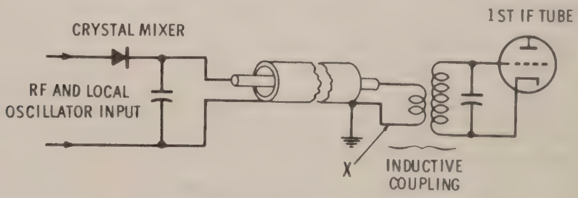


Fig. 6-10. Coupling to IF amplifier.

also are used. A meter, with a bypass capacitor shunting it, is frequently inserted in the circuit at the point marked *X* so that the crystal current may be read. The meter serves as a convenient indicator for adjustment of the level of the local oscillator. Usual crystal currents are something less than 1 milliampere.

DETECTION AND AMPLIFICATION

The method of handling the output of the intermediate-frequency amplifier differs with different types of radar receivers. Detection and subsequent treatment of the signals in a number of types of receivers are described briefly.

Pulse-Type Radar

The transmitter in a pulse-type radar system sends out a series of short energy pulses. These are reflected from any targets in view, and echo pulses are returned to the radar receiver. Accordingly, the receiver listens for echo pulses during the intervals between the transmitted signals. The range of a given target is determined by noting the time interval between transmission of an outgoing pulse and arrival of the corresponding echo from the target in question. The direction of a given target is determined by noting the direction from which the echo pulses arrive. The simplest method of determining target direction involves the use of a highly directional antenna for both transmission and reception. The echo from a given target is then seen only when the antenna is oriented

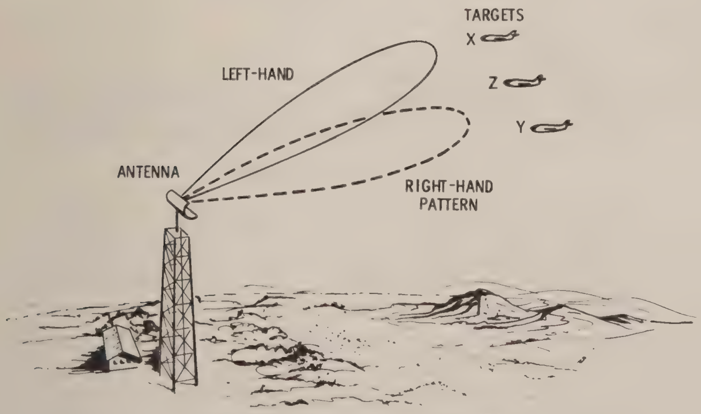


Fig. 6-11. Lobe-switching pattern.

so that the target lies in the beam. A more complicated method of determining target direction in a single plane involves rapid shifting of the radiation pattern back and forth through a small angle. Fig. 6-11 shows the operation of such a system, known as a *lobe-switching* system. The radar alternately transmits and receives with first the left- and right-hand patterns, the pattern being switched several times each second. Obviously, a target at *X* will return a stronger echo when the radar is operating on the left-hand pattern than when it operates on the right-hand pattern. For a target at *Y*, the situation is reversed, and the stronger echo is returned during operation on the right-hand pattern. For a target at *Z*, in the center of the pattern system, an echo of constant strength is returned, since the target is just as far from the center of the left-hand pattern as from the right-hand pattern.

A *conical-scanning* system is used for operation in three dimensions. Here, the beam is swept around the surface of a narrow cone, as shown in Fig. 6-12. A target on the axis of the cone returns an echo of constant strength. A target that is off the cone axis returns an echo amplitude-modulated at the conical scanning frequency (usually 30 cycles). The amount of modulation shows how far the target lies from the axis of the scanning cone; and the phase of the modulation, relative to the phase of the scanning motion, shows the direction in which the target lies. In both lobe-switching and conical-scanning systems, it is necessary to consider the percentage of modulation rather than the actual modulating signal. A nearby target that is slightly off the center of the pattern will return a strong echo with a moderate amount of modulation. A distant target far off the center of the pattern will return a weak echo width, possibly the same *amount* of modulation but with a higher percentage of modulation. The

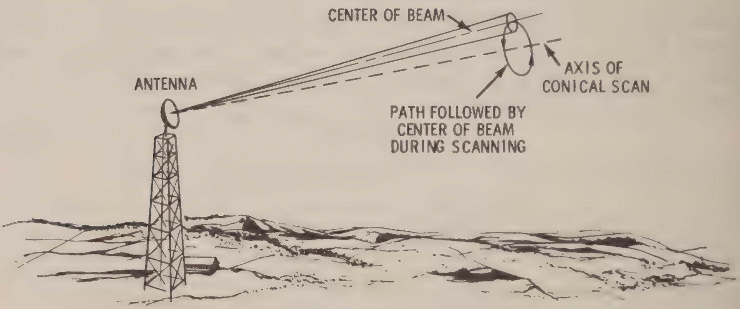


Fig. 6-12. Conical-scan pattern.

modulation itself may be used as an indication of how far the target lies from the center of the pattern, only after the average echo strength has been brought to some standard level.

Search Radar (Pulse-Type)

In a search radar it is desirable to examine all echos. The output of the intermediate-frequency amplifier therefore is detected in a conventional amplitude-modulation detector. Following this, the signal is amplified in a video amplifier and sent to a suitable display system where it may be observed by an operator.

Tracking Radar (Pulse-Type)

The output of the intermediate-frequency amplifier is detected in a tracking radar, as in the case of a search radar, and amplified in a video amplifier to some convenient level. At this point, the desired echo is selected by means of a *gated amplifier* controlled by a range gate. An amplifier of this sort is shown in Fig. 6-13. The tube normally is cut off by the negative bias applied to grid No. 1. At an appropriate time after transmission of the outgoing pulse, the range-gate signal arrives at grid No. 1 and momentarily overcomes the negative bias. The tube does not conduct, however, unless a signal arrives at grid No. 3 during the time that the gate is *opened* by the range-gate signal. Echoes from the target under observation are thus passed by the range gate, and all others are rejected. At the same time, the range of the target under observation is easily determined by noting the time interval

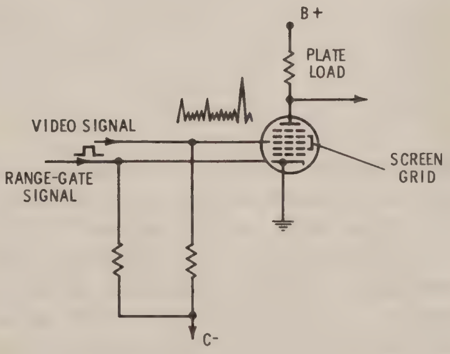


Fig. 6-13. Gated amplifier.

between transmission of the outgoing signal and generation of the range-gate signal. The gated signal, after amplification, is detected again in a low-frequency detector, which recovers the modulation on the echo pulses. The modulation may then be used to control the direction in which the antenna points. (The control is accomplished by means of servo systems.) Either the gated video signal, which contains echoes from the target under observation, or the ungated video signal, which contains all of the echoes received, may be sent to a display system for observation by an operator.

Automatic-Tracking Range Gate

In some radars the range gate is held on the desired echo by an operator; in others, an automatic-tracking range gate is used. In the automatic-tracking system, three gated amplifiers are required. One is controlled by an early gate, one by a signal gate, and one by a late gate. The time relationship of the gates is shown in Fig. 6-14. The early gate opens first; shortly after this the signal gate opens, and finally the late gate opens. The gates then close in the same order.

As shown in Fig. 6-14, there is a slight amount of overlap between the early gate and the late gate. In operation, the signals from the early-gate amplifier and those from the late-gate amplifier are compared. If the group of range gates is set at a range slightly greater than the target range, almost all of the echo will be passed by the early gate and very little will be passed by the late gate. The output of the early-gate amplifier and the resulting difference are used to reduce the range at which the gates are set. If the range is too small, the output of the late-gate amplifier will be greater than that of the early-gate amplifier, and an appropriate correction can be made. The range setting is thus changed automatically until the output of the early-gate amplifier is equal to that of the late-gate amplifier. Under this condition, the output of the signal-gate amplifier contains echoes from the desired

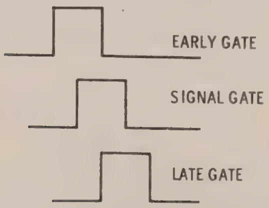


Fig. 6-14. Gates for automatic-tracking range-gate system.

target. As the range changes, the three gates are automatically shifted so that the target continues to appear in the signal gate.

Frequency-Modulated Radar

A specialized type of radar, of which the low-altitude absolute altimeter is a good example, makes use of frequency-modulated transmission rather than pulsed transmission. Consider a transmitter operating in the neighborhood of 500 mc and frequency-modulated in the manner shown in Fig. 6-15.

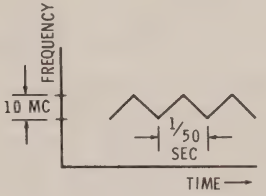


Fig. 6-15. Frequency modulation of FM radar transmitter.

At every instant the transmitter frequency is changing at the rate of 1,000 mc, in one direction or the other. Now, if the ground or any other prominent target, is approximately 500 feet from the transmitter, the echo will arrive after a time interval of 1 microsecond. During most of the time (all of the time, except just before and just after the direction of the frequency change is reversed), the frequency of the received signal will differ from that of the transmitted signal by 1,000 cycles. The range to the target, or the distance to the ground, then can be determined by measuring the frequency difference. This can be done directly by using the transmitter as the local oscillator in the receiver. The intermediate frequency then will be the desired frequency; the difference in frequency between the transmitted and the received signals will be equal to the intermediate frequency. In practice, however, this is difficult, because of the low level of the received signal.

One alternative is the scheme shown in Fig. 6-16. The received signal is picked up on an antenna which is shielded from the transmitting antenna. The signal then is sent to one of a pair of mixers, the other mixer being fed by a low-level signal from the transmitter. Both mixers make use of the same local oscillator, and both input frequencies are shifted by the same amount. The two mixer outputs, which are amplified in identical intermediate-frequency amplifiers, therefore have the same frequency difference as that which exists be-

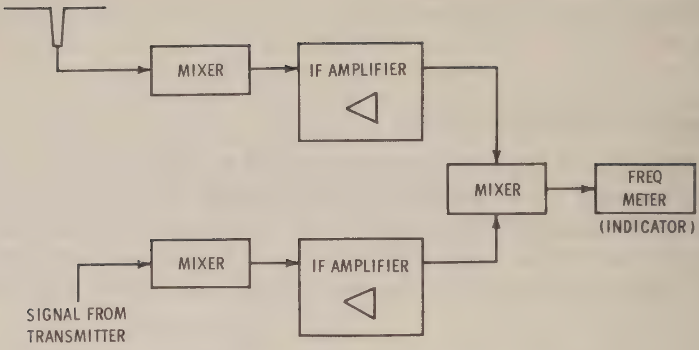


Fig. 6-16. Block diagram of an FM radar receiver.

tween the transmitted and the received signals. After amplification to a suitable level, the two signals from the intermediate-frequency amplifiers are sent to a third mixer. The difference frequency is selected at the output of this mixer by a low-pass filter and is sent to a frequency meter which reads directly in range. Frequency-modulated radar of this sort is useful when only one target at a time is in the combined transmission-reception pattern. Although the example given is a VHF system, the same technique can be used at microwave frequencies.

Continuous-Wave Radar

Another type of radar, CW (continuous-wave) radar, makes use of receiving equipment similar to that used with FM radar. The transmitter, however, is not modulated at all, instead it transmits a continuous signal at constant amplitude and frequency. Suppose, for example, that there is a target at a range of 1,000 wavelengths of the transmitted signal. The received signal having traveled a distance of 2,000 wavelengths, will lag the transmitted signal with a phase difference of 2,000 cycles. Now, if the target is moved toward the transmitter to a distance of 999 wavelengths, the phase difference will be reduced to 1,998 cycles. Suppose, now, that the target moves toward the transmitter at a rate of 1 wavelength per second. The phase difference then decreases at the rate of 2 cycles. If the transmitted signal goes through f cycles in any second (where f is the transmission frequency), the received signal must go through $(f + 2)$ cycles in the same second in order to reduce the phase difference by 2 cycles during the

1-second time interval. The velocity of the target toward the transmitter, or away from it, is given directly by the frequency meter, a frequency difference of 2 cycles that corresponds to a velocity of 1 wavelength per second.

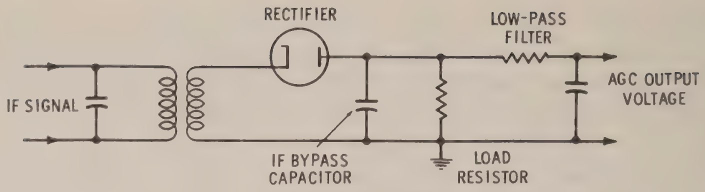
One difficulty with the system as described is the impossibility of telling whether the target is approaching or going away. This may be solved by heterodyning the transmitter signal as it comes from its IF amplifier. This is done with a 1,000-cycle signal. The resulting difference frequency is, then, 1,000 cycles lower than the intermediate frequency which corresponds to the transmitted signal. If this difference-frequency signal is sent to the final mixer in place of the output of the transmitter signal IF amplifier, the frequency meter will show a frequency of 1,000 cycles when the target is stationary. A velocity of 1 wavelength per second toward the transmitter then will be shown by a frequency reading of 1,002 cycles and a velocity of 1 wavelength per second away from the transmitter will be shown by a reading of 998 cycles. Note that the system shows range gate only. If the target moves along a circle the center of which is at the transmitter, the range will remain constant and the system will report no target motion.

AUTOMATIC GAIN AND FREQUENCY CONTROL

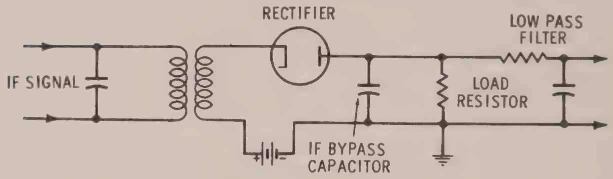
An AGC (automatic gain-control) system in a conventional communications receiver is designed so as to maintain a fixed average signal-level at the output of the intermediate-frequency amplifier, even though the signal level at the reciever input varies over a wide range. Simple control circuits of this sort are seldom used in radar receivers. However, most radar receivers do have some form of AGC in conjunction with AFC (automatic-frequency control) which keeps the receiver tuned to a selected incoming signal.

Conventional Automatic Gain-Control System

The ordinary AGC system used in communications receivers is very simple. It involves nothing more than a detector, fed by the output of the IF amplifier and a low-pass filter (Fig. 6-17A). A portion of the amplifier output is fed to a conventional diode. This rectifies the IF signal and develops a negative voltage across the load resistor that is proportional to the amplitude of the signal. The IF bypass capacitor passes the IF component in the rectifier output, and the voltage



(A) Simple AGC.



(B) Delayed AGC.

Fig. 6-17. Automatic-gain-control systems.

across the load resistor is nothing more than the envelope of the IF signal. This voltage is then passed through a low-pass filter that removes any variation caused by amplitude modulation at audio or higher frequencies. The output of the filter, with the diode connection as shown in Fig. 6-17A, is a negative voltage that is proportional to the average signal level at the output of the IF amplifier. This negative voltage is used to reduce the gain of the IF amplifier, usually by shifting the grid bias of each tube in the negative direction. Now, suppose that an input signal level of 10 microvolts develops a signal of 1 volt at the output of the IF amplifier. What increase in the input level is required to raise the IF output level to 1.1 volts? With 1 volt of AGC voltage, the receiver gain from the receiver input to the IF amplifier output is 100,000. This, of course, is the ratio of 1 to 10 microvolts. With an AGC voltage of 1.1 volts, the receiver gain may be reduced by a factor of 10. The new gain then will be only 10,000 and the required input signal will be 1/10,000 of 1.1 volts, or 110 microvolts. With the system as described, there is only a 10% change in the IF output voltage, but the input signal level is made 11 times as great (a change of 1,000%). In many cases it is desirable to retain the maximum receiver gain for small input signals. The simple AGC system is then replaced by a *delayed* AGC system, such as that shown in Fig. 6-17B. Here, a positive delay voltage is applied to the diode in series with the IF output. This prevents the diode from conducting when the IF signal level is low. Therefore there is no AGC voltage developed until the IF signal level exceeds the bias voltage. Simple AGC systems of this type are not suitable for use in radar receivers, except in rare cases where it is desirable to have the strongest echo control the receiver gain. Ordinarily this is undesirable since the receiver gain is then reduced to bring the strongest echo to some standard level. Other echoes, in this case, are at lower levels and may be overlooked.

Gated Automatic Gain-Control System

It is customary, in tracking radars, to use a form of automatic gain control called gated AGC. The echo passed by the range gate is sent to a rectifier, and the resulting voltage is passed through a low-pass filter, as in a conventional AGC system, in order to remove the error modulation, which results from error in the position of the radar antenna. This system differs from a simple AGC system only in the fact that the AGC voltage is developed from a *single* series of echoes—those from the target under observation. As explained, the gated echo must be brought to some standard level in order to have any fixed relation between the amplitude of the error modulation and the extent of the antenna position error. This is carried out by the gated AGC system.

Sensitivity Time Control System

Although it is not feasible to use an AGC system with a search radar receiver (because the strongest echo received would control the gain of the receiver), it is customary to employ an STC (sensitivity time control) system. This system varies the receiver gain in such a manner that echoes from distant targets are amplified more than those from nearby targets.

For the moment, consider two similar targets, one at a range of 10 miles, the other at twice that range. Since the power level of the transmitted signal falls off as the square of the range, the power delivered to the nearer target is four times that delivered to the farther target. The nearer target, being excited at a power level four times as great as that which excites the farther target, returns an echo that is four times as powerful as the echo returned by the farther target. The echo from the nearer target, traveling only half as far as that from the farther target, suffers only one-fourth as much attenuation in traveling to the receiver. The result is that the echo from the nearer target is received at a power level 16 times as great as that of the echo from the farther target. The power level is proportional to the square of the amplitude, and the echo from the nearer target is at a voltage level four times as great as the voltage level of the echo from the farther target. Both echoes will be displayed at the same amplitude, if the echo from the farther target is amplified four times as much as the echo from the nearer target. This is particularly desirable, since the nearer target otherwise would be displayed as a strong echo that might easily distract the operator and cause him to overlook the echo from the farther target. Uniform signals will be displayed for targets of a given type at all useful ranges, if the receiver gain is made to increase with time in the proper way. The required variation in gain is provided by an STC system. The heart of such a system is a circuit that develops an increasing voltage, which is used to offset a negative grid bias in several IF amplifier stages. The gain-control voltage is reset to zero each time the transmitter is pulsed; it increases during the intervals between successive transmitter pulses. Since increasing the gain-control voltage increases the gain of the IF amplifier, the receiver operates at low gain immediately after the transmitter pulse, when strong echoes may be received from nearby targets. During the listening period, the receiver gain increases so that echoes from distant targets are amplified more than those that arrive earlier from nearby targets. Fig. 6-18A shows the input signal level for echoes from similar targets at different ranges. Fig. 6-18B shows the way in which the

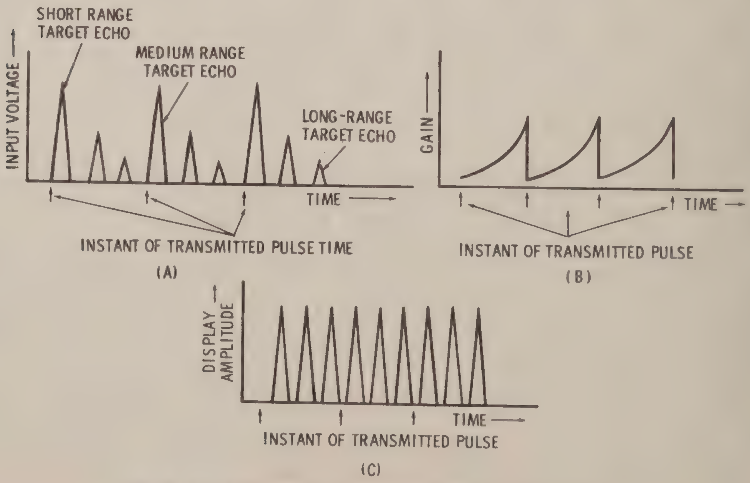

SHORT RANGE TARGET ECHO
MEDIUM RANGE TARGET ECHO
LONG-RANGE TARGET ECHO
INPUT VOLTAGE
TIME
INSTANT OF TRANSMITTED PULSE TIME
(A)
GAIN
TIME
INSTANT OF TRANSMITTED PULSE
(B)
DISPLAY AMPLITUDE
TIME
INSTANT OF TRANSMITTED PULSE
(C)

Fig. 6-18. Operation of sensitivity-time-control system.

receiver gain is varied by the STC system. The amplitude of the echo display is shown in Fig. 6-18C.

Simple AFC System

Most microwave radar receivers incorporate some sort of AFC in order to keep the receiver properly tuned. This is rather simple at microwave frequencies because of the ease with which the frequency of a reflex klystron, used as the local oscillator, can be varied by changing the voltage on the reflector. An AFC system makes use of a conventional FM discriminator tuned to the intermediate frequency. This is fed by the output of the IF amplifier and develops an output voltage of one sign when the mixer output signal is above the intermediate frequency, and a voltage of the opposite sign when the mixer output signal is below the intermediate frequency. The output of the discriminator is, of course, a series of pulses (each echo develops a pulse at the discriminator) but it can be filtered easily to provide a DC signal.

Suppose, now, that the local oscillator is operated at a frequency above that of the incoming signals. If the frequency of the local oscillator is lower than it should be—too close to the frequency of the incoming signal—the output of the mixer will be at a frequency lower than the intermediate frequency. The discriminator, if correctly polarized, will then develop a negative voltage that can be added to the fixed reflector voltage to raise the frequency of the local oscillator. If the local oscillator is operated at a frequency below that of the incoming signals, rather than above, too low an oscillator frequency will develop a mixer output which is above the intermediate frequency. This is due to the fact that the frequency difference between the local-oscillator frequency and that of the incoming signal is too large when the local oscillator frequency is too low. In this case, it is necessary for the discriminator to develop a negative frequency-control voltage when the mixer output is above the intermediate frequency. In either case, when the receiver has been tuned to an incoming signal, it is held in tune (even though the frequency of the incoming signal may change) by the AFC system.

In order to appreciate the merit of an AFC system, assume that a receiver is tuned properly to an incoming signal and suppose that the frequency of the incoming signal changes sufficiently to develop a discriminator output of 10 volts. This mixer output may be developed, for example, by a frequency error of 1 mc in the mixer output. The 10-volt signal from the discriminator will, however, shift the local oscillator frequency by about 10 mc so that the frequency of the incoming signal must have shifted by 11 mc. Therefore, 91% of the frequency shift is compensated for by a change in the local oscillator frequency. A change of 1 mc in the frequency of the mixer output is relatively unimportant, since the IF amplifier has a moderately broad bandpass. A change of 11 mc, however, which would have occurred in the absence of the AFC system, would shift the mixer output entirely out of the bandpass of the IF amplifier and would necessitate manual retuning of the receiver.

Sweep-and-Lock AFC System

With the simple AFC system described it is necessary to tune the receiver to the incoming signals before the AFC system will operate. A more complicated system is able to search for incoming signals and then, when these have been located, it is able to hold the receiver in proper tune. Consider the circuit shown in Fig. 6-19. To be specific, assume that the local oscillator is operated at a frequency above that of the incoming signals, so that an increase in the local-oscillator frequency results in an increase in the frequency of the mixer output signal. Now, consider the circuit which delivers a negative reflector voltage to the local oscillator that is assumed to be a reflex klystron. The circuit consists of thyratron V1, pentode V2, and diode V3. Resistor R1, and capacitor C1, together with thyratron V1, form a relaxation oscillator. Resistors R2 and R3 form a voltage divider to which the cathodes of thyratron V1 and pentode V2 are returned. This supplies bias for both tubes by making the cathodes more positive than the negative supply to which the

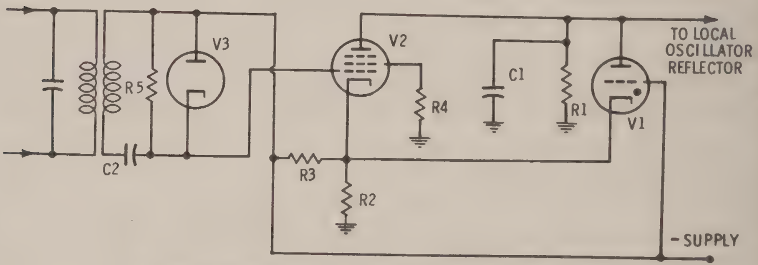


Fig. 6-19. Schematic of a sweep-and-lock AFC system.

control grids are returned. Resistor R4 is nothing more than a screen supply resistor for pentode V2. The tuned circuit is fed by a limiter connected to the output of the IF amplifier. The circuit itself is sharply tuned to the intermediate frequency. Because of the limiter, which is not shown, the output of the tuned circuit (which is fed to diode V3) depends only on the frequency of the mixer output signals and *not* on their amplitude. The output of the tuned circuit is rectified in diode V3 and develops a positive voltage across resistor R5, which is shunted by bypass capacitor C2.

Now consider what happens when there is no incoming signal. As soon as the negative supply delivers its normal voltage, thyratron V1 will fire, charging capacitor C1 and placing a negative voltage on the reflector of the local oscillator. When capacitor C1 has been charged sufficiently to drop the anode voltage of thyratron V1 to a point near the cathode voltage, the tube stops firing. Capacitor C1 then discharges through resistor R1, and the local oscillator reflector voltage as well as the anode voltage of thyratron V1 approaches ground potential. When capacitor C1 has discharged sufficiently to raise the anode of thyratron V1 to the firing point, pentode V2 fires again and recharges C1 once more. A sawtooth voltage, therefore, is applied to the local oscillator reflector, and the local oscillator is swept through a band of frequencies during the intervals between firing of thyratron V1. The waveform of the local-oscillator reflector voltage in the absence of an incoming signal is shown on the left side of Fig. 6-20A, and the frequency is shown on the left side in B.

Assume now, that there is an incoming signal. When the local oscillator reaches a frequency that permits reception of

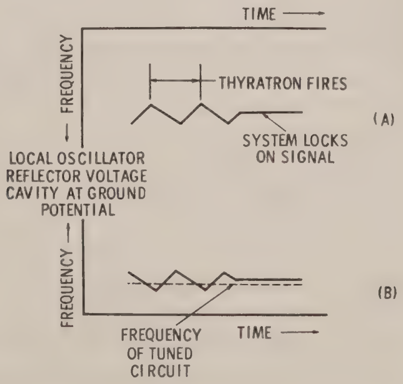

TIME
FREQUENCY
THYRATRON FIRES
(A)
SYSTEM LOCKS ON SIGNAL
LOCAL OSCILLATOR REFLECTOR VOLTAGE CAVITY AT GROUND POTENTIAL
FREQUENCY
(B)
FREQUENCY OF TUNED CIRCUIT
TIME

Fig. 6-20. Local oscillator reflector voltage and frequency.

the incoming signal, there is an output from the IF amplifier. This output is rectified in diode V3 (Fig. 6-19) and develops a positive voltage across resistor R5 which causes pentode V2 to conduct. The additional current through pentode V2 produces a drop across resistor R1, and prevents further discharge of capacitor C1 so that the local-oscillator reflector voltage, as well as the frequency of the local oscillator, are held constant. The system stabilizes itself at a point for which the mixer output is slightly above the frequency of the tuned circuit. This is shown on the right sides of A and B in Fig. 6-20. If, subsequently, the frequency of the incoming signal increases, its frequency approaches that of the local oscillator more closely, and the frequency of the mixer output drops. The frequency of the IF output, therefore, approaches that of the tuned circuit more closely, and the signal delivered to diode V3 (Fig. 6-19) is increased. This makes the grid of pentode V2 more positive and increases the drop across resistor R1. The local-oscillator reflector is then made more negative, and the original increase in the incoming-signal frequency is accompanied by an increase in the local-oscillator frequency. By a similar operation, the local-oscillator frequency is lowered when that of the incoming signal is made lower. The system, therefore, behaves like a simple AFC system when once it has located an incoming signal. It has the advantage, however, that it need not be tuned to the incoming signal to begin with. Operation can be improved slightly if the simple tuned circuit and rectifier are replaced by a discriminator so that either a positive or a negative voltage can be delivered to the control grid of V2.

COMMUNICATIONS RECEIVERS

It was pointed out in Chapter 1 that one of the advantages of operation at microwave frequencies is the great bandwidth that can be obtained with relatively little difficulty. In pulse-type equipment (conventional radar, for example), increased bandwidth permits the use of shorter pulses, as stated in the discussion of intermediate-frequency amplifiers. With shorter pulses, radar equipment is able to make more precise range measurements, since there is less uncertainty about the exact time of arrival of an echo pulse. In addition, with shorter pulses, range resolution is improved; less separation is required between two targets in order to make them easily distinguishable as two separate targets rather than a single large target. In communications equipment where range precision and range resolution are unimportant, increased bandwidth permits an increase in traffic-handling ability. The two fundamental methods of multichannel operation are frequency and time sharing. The discussion of these methods does not apply to any specific types of equipment but is intended instead as a description of the basic techniques involved in multichannel operation.

Frequency Sharing

As an example of frequency-sharing technique, a hypothetical 400-channel system is described. It is assumed that each channel carries a single telephone conversation, and that a frequency range of 250 to 4,500 cps is sufficient for this purpose. Roughly, the 400 channels are separated into 20 groups of 20 channels each. In each group, the 20 signals, each having a frequency range from 250 to 4,500 cps, are combined into a single signal having a frequency range from 250 to 99,500 cps. The 20 group signals, each having this frequency range, then are combined into a single signal having a frequency range from 250 cps to 1,999,500 cps.

Channel Combination—Channel combination involves taking a group of 20 signals, each having a frequency range from 250 to 4,500 cps, and transforming them to a single signal having a frequency range from 250 to 99,500 cps. The signal from each channel is passed through a bandpass filter in order to insure that its frequency range does not exceed the allowable limits of 250 to 4,500 cps. The signal from the first channel is then left unchanged. The signal from the second channel is shifted upward in frequency by 5,000 cps. The signal from the third channel is shifted upward in frequency by 10,000 cps, etc., to the signal from the twentieth channel which suffers a frequency shift of 95,000 cps. The frequency-shifting mechanism for the third channel is shown in Fig. 6-21. A single oscillator supplies a signal, rich in harmonics,

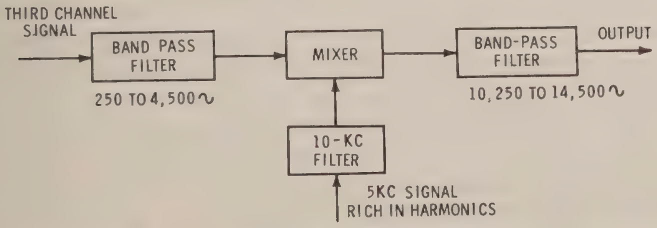


Fig. 6-21. Frequency shifter for third channel.

having a fundamental frequency of 5,000 cps. By means of a filter the second harmonic of this is selected in order to provide a signal at 10,000 cps, the amount of the desired frequency shift. The signal from the third channel, after passing through the bandpass filter, is heterodyned with the 10,000 cps signals in a conventional mixer. The filter at the output of the mixer selects the signal in the band from 10,250 to 14,500 and this is, of course, the sum signal the frequency of which is equal to the sum of the two input signal frequencies. The signals from the 20 channels, therefore, emerge from the bandpass filters, following the mixers with the frequencies shown in the table below. The 20 signals, each lying in a different frequency range, are then combined in a simple network to produce a single signal having a frequency range from 250 to 99,500 cps. Since no two signals can have the same frequency, after frequency shifting, it is impossible for the several signals to interfere with each other. Initially, as shown in Table 6-1, all signals have the same frequency range of 250 to 4,500 cps.

Table 6-1. Frequency Range of Signals

Channel	Frequency (cps)
1	250- 4,500
2	5,250- 9,500
3	10,250-14,500
4	15,250-19,500
20	15,250-99,500

Group Combinations—The 20 groups, each containing the signals from 20 channels, are combined in the manner outlined. The only difference between the channel combination the frequency shift. While the separate channel signals are system and the group combination system is the variation in shifted by multiples of 100,000 cps. The final result, after combination of the groups, is a single signal lying in the range from 250 to 1,999,500 cps, or essentially from 0 to 2 mc. This combined signal, which contains all of the information in the 400 separate channel signals then is used to modulate a transmitter in any convenient way.

Relay Service—In order to relay the combined 400-channel signal, it is not necessary to carry out any separation. At the relay receiver, the incoming signal is heterodyned to a convenient intermediate frequency and amplified in a conventional intermediate-frequency amplifier, having a bandwidth in excess of 4 mc per second. It is then detected in order to obtain the combined signal (0 to 2 mc), which is used without separation of the channels to modulate the relay transmitter.

Signal Separation—At the terminal receiver of the microwave communications system, it is necessary to separate the combined signal to 400 separate signals. This is done easily following detection of the output of the intermediate-frequency amplifier. The 20 groups are separated by bandpass filters, and the 19 groups that suffered frequency shifts before transmission are shifted downward in frequency by appropriate amounts. Each group signal is then once more in the range below 100 kc. Each group then is separated into its constituent 20 signals and 19 of these are frequency shifted in the downward direction by appropriate amounts to recover the original input signals to the 400-channel system.

The basis of a frequency-sharing system is simple—each signal is transmitted all of the time, but it occupies only a small portion of the available frequency range. In the example just described, the available frequency range extends from 0 to 2,000 kc. This is sliced into 400 sections, each one of which is 5 kc wide, and one channel is allocated to each section. Note that each channel occupies almost all of the section allocated to it, but not *quite* all. There is, therefore, some separation between channels, and the required filtering is not impossible.

Time Sharing

As might be expected, the basis of a time-sharing system is that each channel uses the equipment for only a small portion of the time, but the entire available frequency range is used for the signal in each channel during the time when it is transmitted. As an example of a time-sharing system consider an installation that handles 1,000 teletype channels. Teletype information is transmitted in the form of pulses, and it is assumed for this discussion that the pulse rate for a single teletype channel is 100 pulses per second. The signal in a given channel therefore consists of pulses at intervals of one one-hundredth second, and information is transmitted by omitting some of the pulses and transmitting others. With an available frequency range of 2 mc (an IF bandwidth of 4 mc) as in the system above, these pulses may be as short as 1 microsecond. Since the interval between pulses is one one-hundredth second, or 10,000 microseconds, it is relatively simple to transmit a synchronizing pulse that is followed by a pulse from each of the 1,000 channels, if one is to be sent. At the conclusion of this group of pulses, another synchronizing pulse is sent, followed by 1,000 additional signal pulses, and so on.

The great difficulty in time sharing multichannel operation is the problem of sorting out the signals at the receiving end of the system. It is relatively easy to scan 1,000 teletype channels in one one-hundredth of a second, since this allows a 10-μsec interval for examination of each channel. The transmitting system then examines each channel in turn, sending out a pulse if one is called for by the signal in the channel under examination. The problem arises when the received pulse at the other end of the system must be switched to the appropriate output channel of the 1,000 possible ones. To overcome this difficulty, various switching methods are available. However, it often is simpler to make use of frequency sharing. For example, the 2 mc frequency range can be divided by the method described above into 1,000 slices, each one being 2,000 cps wide, and each section then can be occupied by a 1,000 cps signal which is keyed by the teletype. This method calls for complicated filtering, since the harmonics of each signal, after frequency shifting, are likely to interfere with other channels. In practice, a combination of time-sharing and frequency-sharing generally is used.

Mixed Systems

As an example of a system that used both frequency and time sharing, suppose that the available frequency band is divided, on a frequency-sharing basis, into 500 sections. One of these may be reserved for a switching signal, and the others then can carry two teletype channels each. During the second half of the interval, a signal is sent on the switching channel and, at the same time, the other 499 signal channels are connected to the transmitter. At the receiver, one set of 499 channels is connected when the switching pulse is absent and the other set is connected when the switching pulse is present. At first glance this appears to offer little advantage over the straightforward frequency-sharing system. It does, however, simplify the filtering problem to some extent, since the frequency separation is doubled. Carrying the same scheme further, consider a system in which there are 125 frequency-shared channels. Three of these may be reserved for switching signals, and the remaining 122 may each carry 8 teletype channels—a total of 976 teletype channels. This, of course, simplifies the filtering problem tremendously, since only 125 frequencies are involved, and these are distributed over the range from 0 to 2 mc per frequency with only 3 switching frequencies through the use of the system given in Table 6-2.

Table 6-2. Mixed Systems

Channel No.	Switch Signals		
1	Off	Off	Off
2	Off	Off	On
3	Off	On	Off
4	Off	On	On
5	On	Off	Off
6	On	Off	On
7	On	On	Off
8	On	On	On

SPECIAL ASPECTS OF COMMUNICATIONS RECEIVERS

Communications receivers for microwave operation are similar to microwave radar receivers. The receivers themselves are somewhat simpler, in general, but the associated equipment is usually more complex. This results from the fact that channel separation equipment for multichannel operation involves many more circuits than those associated with a radar receiver, for range measurements and automatic tracking. It is usually possible to use a conventional AGC system instead of the gated AGC or STC system required for radar operation. In addition, there is no requirement for the sweep-

and-lock AFC system normally used in radar receivers; a simple AFC system is almost always adequate.

Preselection

It is often desirable to provide some sort of frequency selection before the mixer of a communications receiver, because it frequently is necessary to operate the receiver close to a transmitter operating on a nearby frequency. The situation is worse than in the case of a radar installation, for the radar receiver operates only during the intervals between transmitter pulses, whereas the communications receiver, in a two-way point-to-point installation, operates continuously and at the same time as the associated transmitter. At microwave frequencies, two stages of preselection often are used. In X-band, as well as in K-band operation, where RF amplifiers are not yet available, the premixer frequency selection takes the form of narrow-band microwave filters. These may be resonant cavities tuned to the operating frequency and inserted in the line. Another method is the use of short sections of waveguide having dimensions different from the main transmission system. These are inserted in the line together with tuning stubs and the resulting tuned system then is badly matched for frequencies other than the desired one. Fig. 6-22 shows a single section of narrow waveguide in a normal line for the purpose of increasing the RF selectivity. If a particular frequency is to be rejected, this can be accomplished by a

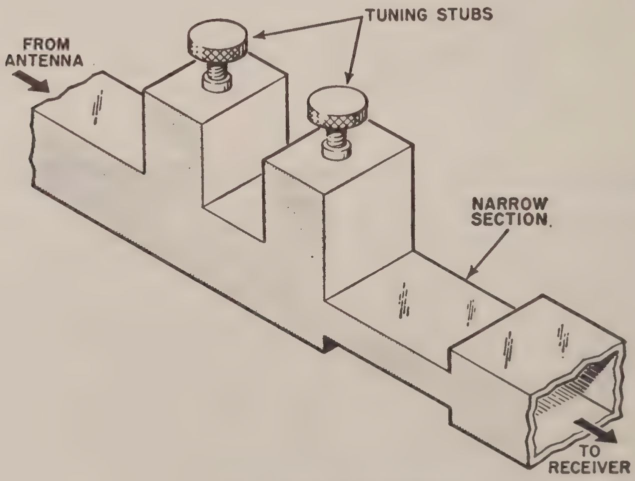


Fig. 6-22. Microwave filter.

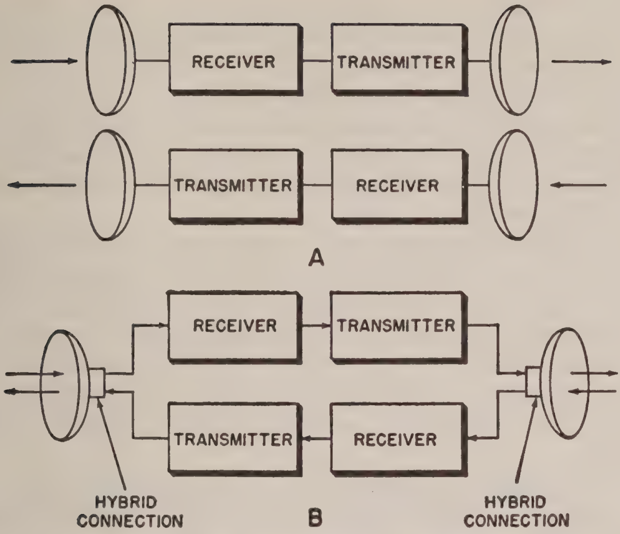

RECEIVER
TRANSMITTER
TRANSMITTER
RECEIVER
A
RECEIVER
TRANSMITTER
TRANSMITTER
RECEIVER
HYBRID CONNECTION
B
HYBRID CONNECTION

Fig. 6-23. Use of hybrid connection.

series of tuning stubs. One of the stubs is adjusted to place a short across the waveguide at the rejection frequency. It then provides complete reflection at that frequency. The remaining stubs are used to eliminate reflection at the operating frequency.

Hybrid Operation

A two-way relay installation normally uses four antennas, as in A of Fig. 6-23. It is desirable, in some cases, to operate with only two antennas instead, as B. Here, it is necessary

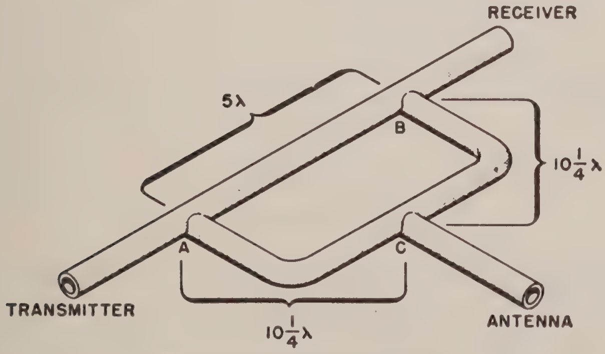

RECEIVER
5λ
B
10$\frac{1}{4}$λ
A
C
TRANSMITTER
ANTENNA
10$\frac{1}{4}$λ

Fig. 6-24. Hybrid connection.

to use hybrid connections at each antenna so that none of the transmitter output will go to the receiver that is connected to the same antenna. A typical hybrid connection is shown in Fig. 6-24. The actual dimensions are not important and as soon as the operation is understood it will be clear that almost any dimensions are feasible so long as distance A-B is an exact number of wavelengths plus ¼-wavelength (although A-C and B-C need not be equal).

Consider the path between the transmitter and the antenna. Signals can travel directly from *A* to *C* and they also can travel by way of *B*. The length of the shorter path is 10.25 wavelengths, and that of the longer path is 15.25 wavelengths. Signals from *A* therefore arrive at *C* in phase with each other, since one path is 5 wavelengths longer than the other. The same thing is true of the path between the antenna and the receiver.

The direct path between *A* and *B*, however, is 5 wavelengths long, and the path by way of *C* is 20.5 wavelengths long. The difference in path length is, therefore 15.5 wavelengths, and signals arriving at *B* by way of the two paths cancel each other, because they have a phase difference of 180°.

A hybrid connection permits a transmitter and a receiver to use the same antenna at the same time without interference. The only difficulty is that some of the received energy finds its way to the transmitter where it is wasted. Under certain conditions, however, it is permissible to waste a portion of the received signal in order to obtain single-antenna operation. This is always the case if the single antenna has more than twice the area of the separate receiving antenna which would otherwise be used.

Chapter 7

Microwave Antennas

The basic problem involved in terminating a transmission line is that of impedance matching. When the impedance of the load differs from the characteristic impedance of the transmission line, a portion of the energy sent to the load is reflected at the line-load junction. It is possible, of course, to cancel the reflected signal through the use of tuning stubs, or equivalent devices, so long as the load impedance has a resistive component which is neither zero nor infinitive.

TRANSMISSION LINE-TO-LOAD IMPEDANCE MATCHING

The system shown in Fig. 7-1 consists of a transmission line, a load, the impedance of which does not match that of the line, and a tuning device. When the tuning device is adjusted properly, section P of the transmission line (between the generator and the tuning device) is flat; that is, there is no reflected signal on this section of the transmission line.

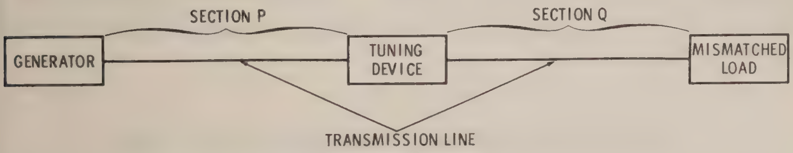


Fig. 7-1. Line-to-load matching.

The tuning device, however, has no effect on the reflected signal present on section Q (between the tuning device and load). Thus, even after the reflected signal on section P has been cancelled by the tuning device, there is a residual impedance mismatch loss because of the presence of the reflected signal on section Q. This residual loss can be reduced by placing the tuning device close to the load and thus shortening section Q. It can also be reduced by making the load impedance approximately equal to the characteristic impedance of the line so that the amplitude of the reflected signal is decreased. Both of these methods are used in practice. It happens occasionally, however, that the load impedance cannot be made even approximately equal to the line impedance. One example of this is the case of a filled wave guide that is connected as a load on a hollow wave guide. The residual loss then is minimized by using a tapered section, the impedance of which changes gradually. At one end the impedance of the tapered section matches that of the load; at the other end it matches the impedance of the transmission line that feeds the load. Fig. 7-2 shows a section of this type—where section Q is a tapered section of wave guide. Instead of an abrupt change in impedance at the load end of section Q (Fig. 7-2), there is a slight mismatch at *every* point on the tapered section. There is, then, a negligible reflected signal at the load end of section Q and only a small reflected signal at other points.

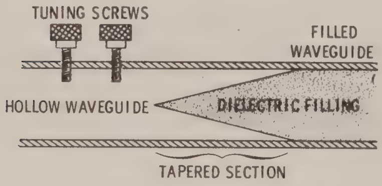


Fig. 7-2. A tapered section of transmission line.

GENERATOR-TO-LINE IMPEDANCE MATCHING

The problem of impedance matching between a generator and a transmission line is essentially the same as the problem of matching between a line and its load. The techniques are the same: tuning devices are placed close to the sending end of the line, and a tapered section sometimes is used if the mismatch is great.

Every antenna system has a part that is used as a coupling of some sort between a transmission line and the antenna. In a transmitting antenna this is a line-to-load problem; in a receiving antenna, it is a generator-to-line problem. Both problems, however, are essentially the same, since both are a question of impedance matching. When a conventional radiator is connected to a conventional transmission line, the problem is easy to visualize. The radiator acts as a coupling element between the line and space, and it is only necessary to adjust the radiator and the line to obtain an impedance match. A center-fed half-wave antenna, for example, has an impedance of about 70 ohms and may be connected directly to a 70-ohm line. In the case of waveguide, however, the problem appears to be more complicated.

It was pointed out in Chapter 3 that the impedance of a transmission line may be given either by the voltage/current ratio or by the ratio of the transverse electric-field strength to the transverse magnetic-field strength. On this basis, a hollow wave guide has a characteristic impedance for each mode. The impedance depends on the dimensions of the guide among other things; the impedance of space is the same as that of a wave guide of infinitely great height and width. Several methods for making the required impedance match are discussed later in the chapter, and only one will be mentioned here. This is the *tapered horn* (Fig. 7-3) which is a simple tapered section of the type shown in Fig. 7-2. A proper impedance match allows energy to be radiated from the end of the wave guide, without reflection back toward the transmitter. It also allows the guide to be excited by external radiation without reflecting a portion of the energy back into space.

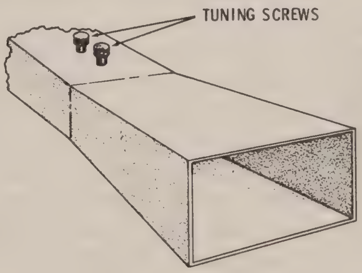


Fig. 7-3. Open-end view of a tapered horn.

MICROWAVE ANTENNAS

In general, a microwave antenna consists of both a focusing and a feed system. For transmission, the feed system couples a wave guide to space and the focusing system collects radiation from the feed system and focuses it in a beam. For reception, the focusing system receives incoming radiation and focuses it on the feed system where it excites a wave guide. Focusing and feed systems are discussed later in this chapter, as well as the use of reflectors and metal lenses for focusing.

DISSIPATIVE TERMINATIONS

It is sometimes desirable to test a microwave transmitter by feeding a load without radiating a signal. A dissipative termination is used as a load in place of a transmitting antenna. As its name implies, a dissipative termination is merely a load which is able to absorb power and dissipate it in the form of heat. At low power levels, a flat card coated with resistive material may be used. The card is placed across the end of the wave guide or coaxial line and makes conductive contact with the walls of the wave guide or with the outer and inner conductors of the coaxial line. If the conductivity of the coating is chosen correctly, the card will have an impedance, measuring in ohms per square (explained in Chapter 3), equal to the characteristic impedance of the transmission line. The line then will be terminated properly and all the energy that strikes the card will be absorbed; there will be no reflection. If precise matching is important, a tuning device may be installed near the card. This method is adequate at low power levels, but the dissipation of the card is not sufficient to enable it to handle any great amount of power. At higher power levels, a useful termination must distribute the absorption over a fairly large region so that no great amount of heat is developed at any one point. This sometimes is called a *lossy line* termination and takes the form of a wave guide or a rigid coaxial line made of iron instead of brass.

Since iron has a relatively high resistance (among conductors), the IR losses are not negligible, and power is absorbed in the walls of the wave guide or in the conductors of the coaxial line. The lossy line ends in a short-circuit termination. This, of course, produces a reflected signal, but the reflected signal is attenuated (by losses) on its way back toward the transmitter, and if the lossy line is sufficiently long, the magnitude of the reflected signal at the input to the lossy line will be negligible. Still another method, which may be used instead of the iron-conductor method, is the use of a lossy dielectric inside the wave guide or coaxial line. Sand is frequently used, and the resulting terminating system is called a *sand load.* A tapered section of the sort shown in Fig. 7-2 is used to assist in the impedance matching between the normal line and the dielectric-filled line, which is, of course, filled with a *lossy* dielectric. This method and the one just described are frequently combined in the form of a line which has *both* dielectric and conductor losses. For moderate power, the lossy line of either sort is provided with metal fins which enable it to radiate the absorbed power in the form of heat. At higher power levels it is necessary to provide forced-air cooling by means of blowers; for still higher power handling capacity it is necessary to use water cooling.

RADIATION PATTERNS

There are two distinct point sources to be considered in a discussion of radiation patterns. One of these is an ideal point source, sometimes called an *ideal antenna,* which radiates uniformly in all directions. The other is called a *practical* point source, or an *elementary source,* which is the basic element that is considered in calculating the performance of a real antenna.

Ideal Antenna

The ideal antenna cannot be achieved in practice; it would not be desirable, if it could be constructed, since it radiates uniformly in all directions instead of concentrating the radiated energy in a beam. The usefulness of the ideal antenna lies in the ease with which its performance can be calculated. This performance, in turn, can be used as a base on which the performance of a real antenna can be readily calculated. A single example will suffice to make this clear; suppose that a given transmitter delivers 10 kilowatts to a transmitting antenna and that a receiving antenna is located at a point 50,000 feet (slightly less than 10 miles) distant. If the receiving antenna consists of a circular reflector having a radius of 1 foot, how much power is delivered to the wave guide which connects the receiving antenna to the receiver? In the following calculation, the effect of reflection from the earth is neglected.

To begin with, assume that the transmitter feeds an ideal antenna. Imagine a spherical surface having its center at the transmitter and having a radius of 50,000 feet. The sphere has a surface area of $4\pi\ (50{,}000)^2$ square feet, or 10,000,000,000 (π) square feet. For simplicity, 10,000,000,000 may be expressed as $10^{10}(\pi)$ square feet. The 10 kilowatts of transmitted power is radiated uniformly in all directions; consequently, the power falling on the surface of the sphere at *any point* is $10/10{,}000{,}000{,}000(\pi)$ kilowatts per square foot. Since 10 kilowatts are equal to 10,000 watts or 10,000,000,000 microwatts, this reduces to a *power density* (power per uit area) of $1/(\pi)$ microwatts per square foot. The receiving antenna has a circular reflector with a radius of 1 foot. Since the area of a circle is $(\pi)\,r^2$ where r is the radius, this reflector has an area of (π) square feet. The ideal antenna (at a distance of 50,000 feet, and with a power input of 10 kilowatts) delivers $1/(\pi)$ microwatts per square foot at the receiver site and the reflector picks up exactly 1 microwatt. Assuming good focusing and good matching, almost all of the incoming power (1 microwatt) will be delivered to the receiver.

Now, suppose that the transmitter feeds a real antenna which concentrates the radiated energy in a beam. The behavior of such an antenna can be described in terms of its *power gain.* If, for example, the real antenna has a power gain of 10, it will deliver 10 times as much power (along its beam) as an ideal antenna having the same input. It delivers less power than the ideal antenna, of course, in other directions. With the arrangement described, a real antenna having a power gain of G will deliver G microwatts to the receiving antenna. The general expression for the power density delivered by a real antenna at distance D (in feet) is $PG/4(\pi)D^2$ watts per square foot, where P is the power input in watts and G is the power gain. The ratio between the circumference and the diameter of any circle is π which is approxmately equal to 3.14159 or, for ordinary purposes, 3.14.

Elementary Source

An elementary source differs from an ideal source by having an axis. It radiates best in directions perpendicular to the axis and does not radiate at all along the axis. For points lying in a given direction from the elementary source, the signal level is proportional to $1/D$ (where D is the distance from the source to the point of measurement) and the power level is proportional to $1/D^2$. This behavior is similar to that of any other radiating system. In addition, the signal level is zero for points on the axis of the elementary source, and for points off the axis the signal level is proportional to the

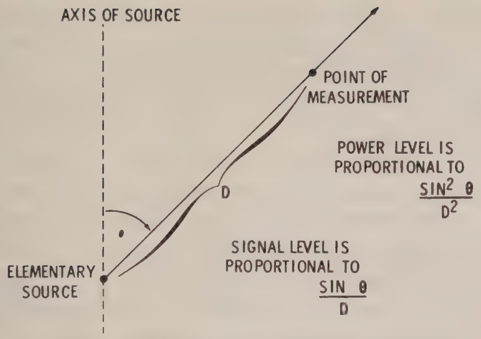
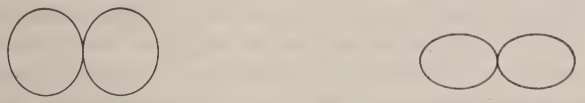

AXIS OF SOURCE
POINT OF MEASUREMENT
POWER LEVEL IS PROPORTIONAL TO $\frac{\sin^2\theta}{D^2}$
D
θ
ELEMENTARY SOURCE
SIGNAL LEVEL IS PROPORTIONAL TO $\frac{\sin\theta}{D}$

Fig. 7-4. Radiation from elementary source.

sine of the angle formed by the axis and the line joining the source to the point of measurement (Fig. 7-4). Now, consider a plane which contains the axis of the elementary source, and choose a point that lies in the plane *and* in a direction, from the source, that is perpendicular to the axis. Let this be a reference point, and let the signal level at this point be chosen as a reference level. The signal levels for other points in the plane, lying at the same distance from the source but in different directions, may then be plotted as shown in Fig. 7-5A. The power level is plotted as in Fig. 7-5B.



(A) Relative signal level. (B) Relative power level.

Fig. 7-5. Radiation pattern of elementary source.

Beamwidth of Elementary Source

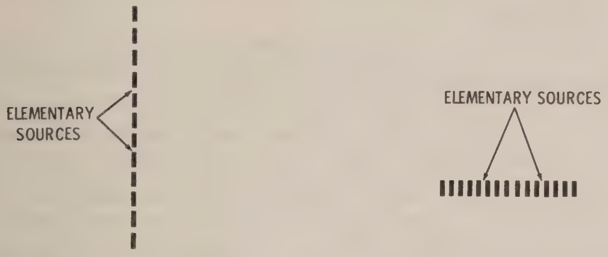
Inspection of Fig. 7-5B reveals that the power level drops to one half of its maximum in directions that are 45° away from the direction of the maximum power level. The *beamwidth,* which is measured between the two half-power directions, is therefore 90° in a plane containing the axis. There is, of course, no beaming in the direction of the axis. Note that the signal level, in the half-power directions, is .707 of its maximum, since .707 is the square root of 0.5.

Power Gain of Elementary Source

Suppose that an ideal antenna delivers a power density of 1 microwatt per square foot at a distance of 1,000 feet. The total power input to the ideal antenna may be found by considering the surface of a sphere having a radius of 1,000 feet. If the ideal antenna is located at the center of this sphere, the power density at any point on the surface will be 1 microwatt per square foot. The surface of the sphere has an area of 4,000,000(π) square feet, and a total power input to the antenna, which is equal to the total radiated power, is then 4,000,000(π) microwatts or 4(π) watts. If an elementary source delivers the same power density (1 microwatt per square foot at a distance of 1,000 feet in a direction perpendicular to its axis where it radiates the strongest signals), the total power input may be found in the same way. As before, the spherical surface has an area of 4,000,000(π) square feet, but the power density is not uniformly equal to 1 microwatt per square foot. In fact, in the direction of the elementary-source axis, the power density is zero. If the total radiated power is computed by adding the power density at each square foot of the surface, it is found that the input to the elementary source is only (8/3) (π) watts—two-thirds as much as the power input to the ideal antenna. Now, if the power input to the elementary source is multiplied by 1.5, which brings it up to 4(π) watts, the power level at any point will also be multiplied by 1.5. It follows then that the power density in a direction perpendicular to the axis at a distance of 1,000 feet will be 1.5 microwatts per square foot. Since the elementary source delivers 1.5 microwatts per square foot (in the direction of maximum radiation) while the ideal antenna delivers only 1 microwatt per square foot with the same power input, the elementary source has a power gain of 1.5.

RADIATION FROM LINE SOURCE

As the next step in considering the behavior of real antennas, consider a line source. This may be thought of as a great many elementary sources in a straight line. As might be expected, the orientation of the elementary sources has some effect on the radiation pattern of the complete line source. That is, one sort of pattern is obtained when the axis of the elementary sources are parallel to the line along which they are placed, and another sort of pattern that is similar to the first but not identical with it is obtained when the axes of the elementary sources are all perpendicular to the line, but parallel to each other. The two arrangements are shown in Figs. 7-6A and B. They may be thought of as end-to-end, and side-by-side arrangements. In order to simplify the picture, it is convenient, at first, to assume that the line source is made up of a large number of ideal point sources, which are not directional. A line source of this type may be called an ideal line source.



(A) End-to-end line source.

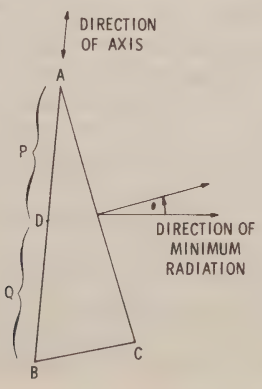
(B) Side-by-side line source.

Fig. 7-6. Elementary sources arranged as a line source.

Ideal Line Source

Like an elementary source, the ideal line source has an axis, which is the direction of the line source itself. The maximum radiation is in directions that are perpendicular to the axis, and the signal strength in any particular direction depends only on the angle between that direction and the axis. In a given direction, of course, the signal level is proportional to 1/D and the power level is proportional to $1/D^2$, where *D* is the distance between the source and the point of measurement. The entire radiation pattern, as in the case of an elementary source, can be figured from the pattern in any plane that contains the axis of the source. This may be calculated with the help of Fig. 7-7. Here, line AB represents the ideal line source; the arrow pointing in direction *X* shows the

Fig. 7-7. Line source.



direction in which the signal level is to be calculated; and AC is a line perpendicular to this direction. It is necessary, of course, to compute the pattern on the basis of signal level rather than on power, since the signals from the various parts of the line source must be added with due regard to any phase differences.

For radiation in a direction perpendicular to the source, all portions of the source contribute signals of the same phase, and these may be added directly. For radiation in other directions, the signals from various portions of the source are not in phase with each other, and when the various signals are added, the sum may vanish. For example, in the case shown in Fig. 7-7, signals from *B* on the source travel farther than signals from *A* in reaching a distant point. The difference in path length is distance BC, which depends on the length of the source and on angle Θ. For radiation in the preferred direction, BC is zero and, as noted, signals from all parts of the source add directly. By the time BC reaches a length of slightly less than one-quarter wavelength, the phase differences between signals from different parts of the source are enough to reduce the signal level to .707 of its maximum value. The half-power direction is, accordingly, one for which BC is a little less than 1 quarter-wavelength. When BC increases to one-half wavelength, the signal level is reduced to .64 of its maximum value and the power is reduced to .41 of its maximum value. When BC is equal to exactly 1 wavelength, the radiation from the line source is zero. This happens because the signal from *B* arrives at a distant point one cycle later than the signal from *A*. The signal from *D*, having an extra distance of 1 half-wavelength to travel, arrives at a distant point precisely out of phase with the signal from *A* and cancels it.

For every point, such as *P* (in one half of the source) there is a point, *Q*, in the other half (where the distance between *P* and *Q* is equal to the distance between *A* and *D*—half the length of the source), and the signal from *Q* cancels the signal from *P*. Thus, the half-power direction is reached when BC is just under one-quarter wavelength, and the null—the direction in which there is *no* radiation—is reached when BC is 1 wavelength. When BC increased to 1.5 wavelengths, the signal is once more at a maximum, but a maximum which is much smaller than the signal in the main beam.

To understand this, examine Fig. 7-8. The line source can be thought of as consisting of three parts, FG, GH, and HI. If distance IK is 1.5 wavelengths, distance HJ will be exactly

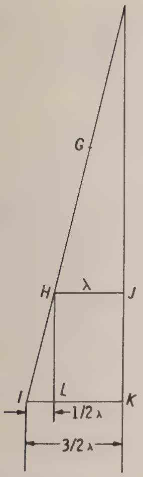
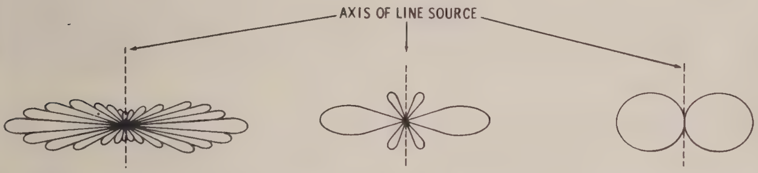


Fig. 7-8. Calculations of first side lobe.

1 wavelength and the signals from FG and GH will cancel completely. There remains the signal from HI, which is at an angle that makes distance IL equal to one-half wavelength. The situation with regard to part HI of the source is exactly the same as the situation for the entire source when distance BC (Fig. 7-7) is one-half wavelength. Under this condition, with BC equal to one-half wavelength, the signal level is .64 of its maximum level. When IK is 1.5 wavelengths, and IL is one-half wavelength, only one third of the source is involved in successful radiation and the signal level is .64/3 or about .21 of its maximum level. The power level in this direction is therefore about .04 of its maximum. The low-level beam is called a side-lobe; there are additional side lobes with signals at decreasing levels whenever distance L is any odd number of half wavelengths. For the directions in which L is an even number of half-wavelengths, there is no radiation. Fig. 7-9A shows the signal-level pattern from a long-line



(A) Long line source.

(B) Shorter line source.

(C) One half-wavelength line source.

Fig. 7-9. Line-source patterns.

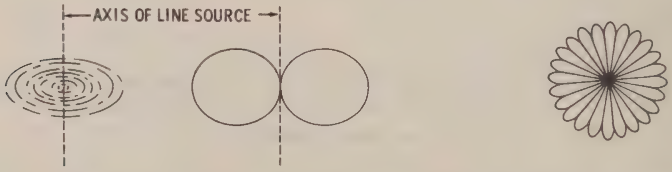
source, and Fig. 7-9B shows the signal level pattern for a shorter source. The pattern for a line source only one-half wavelength long is shown in Fig. 7-9C.

End-to-End Line Source

In the case of an end-to-end line source, the pattern is modified by the characteristics of the elementary sources. Therefore, for radiation in directions close to the source axis, which is the same as the axis of each elementary point source, the signal level is reduced, because each elementary source radiates less in that direction. Thus, in a direction 45° away from the direction of maximum radiation, the signal level from an elementary source is reduced to .707 of its maximum value. The signal level from an end-to-end line source is similarly reduced to .707 of the value it could have, if the source were replaced by an ideal line source radiating the same signal level perpendicular to the source axis. For a moderately long-line source, this reduces the signal level in the smaller side lobes and has only a negligible effect on the main beam signal.

Side-by-Side Line Source

The pattern from a side-by-side line source is similar to that from an end-to-end line source, except that it is no longer uniform in a plane perpendicular to the line-source axis. This results from the fact that such a plane contains the axes of all the elementary sources, as shown in Fig. 7-10. Fig. 7-10A shows a signal-level pattern (main lobe only) of an ideal line source *or* of a moderately long end-to-end line source. Fig. 7-10B shows a side-by-side line source in which the axes of the elementary sources are perpendicular to the paper of this page.



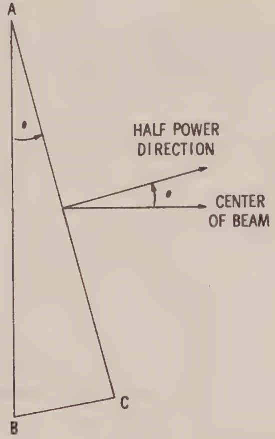
(A) End-to-end sources. (B) Side-by-side sources.

Fig. 7-10. End-to-end and side-by-side patterns.

Beamwidth of Line Source

The beamwidth of a line source several wavelengths long is the same whether it is an ideal line source, an end-to-end line source, or a side-by-side line source. The width can be computed by referring to Fig. 7-11. In the half-power direction, distance BC is about one-quarter wavelength. The sine of the angle, Θ, which is the angle between the center of the beam and the half-power direction, is BC/AB. In the half-power direction, BC may be taken as one-quarter wavelength. Distance AB, which is the length of the line source, may be taken as *N* wavelengths. The sine of Θ then is equal to (¼)/N, or ¼N. The angle, Θ, is then equal to ¼N radians or, taking one radian as equal to 60° (a reasonable approximation), to 15/N°. The beamwidth is, of course, equal to 2Θ, since it is the angle between the two half-power directions. On this basis, the beamwidth may be taken as approximately equal to 30/N° for a line source length *N* wavelengths. This expression is reasonably precise when *N* is greater than 1.

Fig. 7-11. Beamwidth.



Power Gain of Line Source

The power gain of an ideal line source may be taken as equal to 2N, where *N* is the length of the source in wavelengths. If the source is several wavelengths long, as pointed out previously, it will have the same radiation pattern, whether it is an ideal line source or an end-to-end source. If, however, it is a side-by-side line source, the directivity in the plane perpendicular to the axis of the line source will raise the power gain to 3N. This is the product of the line-source gain (2N) and the elementary source gain (1.5).

RADIATION FROM A TWO-DIMENSIONAL SOURCE

The radiation pattern from a plane source is shown in Fig. 7-12 (the side lobes are neglected). This may be obtained by considering the phase source as consisting of a very large number of ideal line sources close together. If the source has a height of *N* wavelengths and a width of *M* wavelengths, the beam height between half-power directions will be about 30/N°, and the beamwidth, also taken between half-power directions, will be about 30/M°. The power gain of the source will be nearly equal to the product of (2N) by (2M), or 4NM. Just as the ideal line source may be replaced by an end-to-end arrangement of elementary sources without altering the pattern or the power gain, so the ideal plane source may be replaced by an arrangement of elementary sources without affecting its performance, provided that the axes of the elementary sources are parallel to each other and parallel to the plane of the plane source. The power gain of the plane source is 4NM, and if all of the power is radiated in *one* direction (instead of two directions as shown in Fig. 7-12), the power gain will be raised to 8NM. NM is the area of the source in square wavelengths so that the power gain is proportional to the source area in square wavelengths. This is the basis of the maximum-range equations developed in Chapter 2.

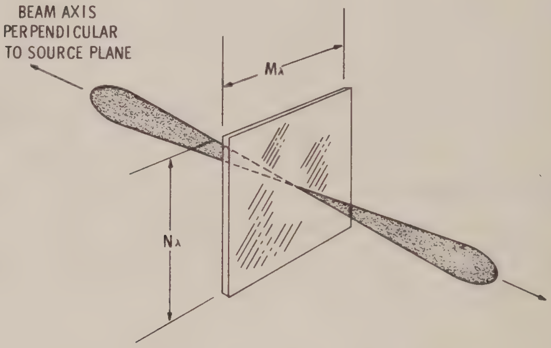


Fig. 7-12. Radiation pattern from plane source.

RADIATION FROM LINE ARRAY

A radiating system which consists of a number of point sources arranged in a line is called a line array, or a linear array. The simplest linear array consists of a pair of point sources which are excited in phase with each other: more complicated linear arrays consist of larger numbers of point source—more than one hundred in some cases.

Pattern of Linear Array

In determining the pattern of a linear array it is necessary to know the number of radiating elements and the spacing between them. When these quantities are known, any linear array can be related to a corresponding line source by considering that the line source has been divided into a number of equal parts, each of which has been replaced by an element of the linear array. Thus, a linear array consisting of two sources spaced 5 wavelengths apart is more or less equivalent to a line source which is 10 wavelengths long. A linear array of four sources with spacing of 2.5 wavelengths is more like the 10-wavelength line source; and an array of 10 sources with spacing of 1 wavelength is much like the corresponding line source. When the spacing between the elements of a linear array is one-half wavelength or less, the radiation pattern of the array is identical with that of the equivalent line source for all practical purposes.

If, for example, a linear array consists of 20 elements with spacing of one-half wavelength, its pattern will be identical with that of a 10-wavelength line source. The radiation pattern will not be changed to any great extent by adding additional sources and, at the same time, reducing the spacing. The pattern will, however, be changed, if additional sources are added without changing the spacing. Two additional sources, for instance, will make the linear array equivalent to a line source of 11 wavelengths. When the spacing between the elements of a linear array is greater than one-half wavelength, the radiation pattern differs from that of the corresponding line source by having more of the energy in the side lobes. As a single example, Fig. 7-13A shows the radiation pattern from two sources spaced 5 wavelengths apart, and Fig. 7-13B shows the pattern from the equivalent 10-wavelength line source. As the number of elements in the linear

Fig. 7-13. Pattern of linear arrays.

array is increased, the outer side lobes become weaker and the central lobe, as well as the adjacent side lobes, becomes stronger. As the number of elements approaches 20, even the side lobes close to the main lobe begin to shrink, and when the linear array consists of 20 elements or more, the pattern is like that in Fig. 7-13B.

Power Gain of Linear Array

When the elements of a linear array are spaced one-half wavelength apart, or when they are closer, the power gain of the array is equal to that of the corresponding line source. This is equal to 2N, where *N* is the length of the line source in wavelengths. The appropriate value of *N* is found by multiplying the number of elements in the linear array by the spacing (in wavelengths) between adjacent elements. When the spacing between the elements of the linear array is greater than one-half wavelength, the power gain is simply equal to the number of elements in the array. The second rule may be verified easily.

Consider, for example, a linear array consisting of four sources with spacing greater than one-half wavelength, and compare it with a single source. Assume that the input impedance of each source is 70 ohms, and that the single source receives an input of 280 watts. Each source of the linear array must then receive an input of 70 watts if the power input to the array is to be equal to 280 watts. The power radiated in a given direction by each source in the linear array is then one fourth as great as the power radiated in the corresponding direction by the single source. At equal distances, the signal level from each of the four sources then will be half as great as the signal level from the single source. Halving the signal level reduces the power level to one fourth, since power is proportional to the square of the signal level.

There are, however, four sources in the linear array. In the main beam of the array, the radiation pattern consists of all the signals added directly. The signal level developed by the array is, therefore, twice as much as the signal level developed by the single source. This is because a 2/1 ratio in signal level is equivalent to a 4/1 ratio in power level. Consequently, the power gain of the array is equal to four, which is the number of elements. If the directivity of the separate sources in the linear array is in the same direction as the directivity of the array, the power gain of the sources may be neglected, as in the case of an end-to-end line source. If, however, the directivity of the separate sources is in a different direction from that of the linear array, the power gain of the array must be multiplied by the power gain of the sources, as in the case of a side-by-side line source.

TWO-DIMENSIONAL ARRAY RADIATION

The radiation pattern of a two-dimensional array bears the same relation to that of a plane source as the pattern of a linear array bears to that of a line source. The array may be considered as consisting of a set of linear arrays. The power gain and pattern of each linear array is computed as described. Next, the linear arrays may be combined in the same way. If the spacing between linear arrays is greater than one-half wavelength, the power gain of the plane array is obtained by multiplying the gain of each linear array by the number of linear arrays. If the spacing between linear arrays is one-half wavelength or less, the power gain of the plane array is obtained by multiplying the power gain of the linear arrays by twice the product of the number of linear arrays and the spacing, in wavelengths, between the linear arrays. The elements in an array are seldom closer than a half wavelength. It is therefore simpler to remember that the power gain of a two-dimensional array is equal to the number of sources involved. This is true, however, only if the spacing is one-half wavelength or more between adjacent sources.

INTERPRETATION OF RADIATION PATTERNS

Although it is customary to speak of the power gain of an antenna system, the radiation pattern normally is plotted in terms of signal level rather than power. The interpretation of radiation-pattern plots is discussed here in relation to communications transmitters, communications receivers, and radars.

Communications Transmitter Patterns

Consider the situation shown in Fig. 7-14. Here the radiation pattern of a transmitter is plotted in terms of signal level (solid line) and power level (dashed line). The relative power level is, of course, equal to the square of the relative signal level. That is, in the direction of *Y*, the signal level is only one-half as great as in the direction of *X* (which lies in the

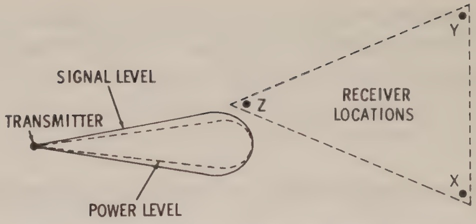


Fig. 7-14. Use of transmitter radiation pattern.

center of the main beam), and the power level in direction *Y* is only one-fourth as great as the power level in direction *X*. Assume that a receiver at *X* delivers a certain output. An identical receiver at *Y*, which is just as far from the transmitter as *X*, will deliver only one-half as much output signal as the receiver at X, or one-fourth the output power. This is shown, of course, by the radiation pattern—the signal-level plot shows a relative level of one-half, whereas the power-level plot shows a relative level of one-fourth. Receiver outputs are usually described in terms of volts rather than watts—signal level rather than power level—and, consequently, the solid line plot (of signal level) is the more useful of the two. A more important reason for the use of the signal-level plot is apparent when consideration is given to a third receiver located at *Z*, which is midway between *Y* and the transmitter. Since the receiver at *Z* is only half as far from the transmitter as the receiver at *Y*, it receives twice the signal strength and four times the power received by the receiver at *Y*. Its output, therefore, is twice as great (in volts) as that of the receiver at *Y* and is equal to that of the receiver at *X*. The merit of the signal-level plot, as opposed to the power-level plot, is now obvious. When the signal level is halved, it is necessary to halve the distance to the receiver in order to obtain the same receiver output. The transmitter radiation pattern, in terms of signal level, may, therefore, be considered as a plot of relative receiver output in volts for receivers in different directions but at equal distances from the transmitter, or it may be considered as a plot of relative receiver distance in different directions for equivalent receiver performance.

Communications Receiver Patterns

The directivity of an antenna system used for reception is the same as its directivity when it is used for transmission. For this reason, the radiation pattern plot may be used also as a reception-pattern plot. As in the case of a transmitting antenna, the pattern is usually plotted in terms of power level. It may be interpreted as the relative receiver output in volts developed by incoming signals of constant strength arriving from different directions. If, for any reason, receiver power output is of greater interest than the voltage output, the pattern may be plotted in terms of relative power level. This, however, is seldom done.

Radar Patterns

The foregoing discussion of communications transmitter and receiver patterns applies equally well to radar beacons, since these operate on a transmitter-receiver basis. For conventional radar operation, however, where a signal is transmitted to a target and an echo is returned, it is necessary to take into account both the radiation and reception patterns. If, as is usual, the same antenna is used for both transmission and reception, the two-way pattern can be obtained by squaring the values obtained for the one-way pattern. Thus, the two-way signal-level plot is identical with the one-way power-level plot for a given antenna. As before, the resulting plot may be considered as a plot of relative receiver output for identical targets at equal distances. If, however, it is desirable to know the ranges (on identical targets) for targets in different directions for some standard receiver output, a one-way signal-level plot must be used instead of a two-way plot. If the signal level has half of its maximum value, a target at half the range of a standard target in the center of the beam will receive the same signal level as the standard target. It then returns an echo of the same strength as the standard target, over only half the distance (which results in twice the echo signal level) to an antenna which is half as effective for reception. (The vertical polar diagrams discussed in Chapter 2 are based on one-way signal-level plots.)

ARRAY METHOD OF ANTENNA DESIGN (USING REFLECTORS)

Design of a microwave antenna system by the array method is a relatively simple task. When the desired radiation pattern has been decided on, the first step in carrying out the design is the specification of a two-dimensional source which has the required radiation pattern. The second step is the design of a radiation (or reception) system which is a satisfactory approximation of the specified two-dimensional source. At frequencies below the microwave region, the resulting radiation system is often a two-dimensional array of separate sources, and the method of design takes its name from systems of this sort. At microwave frequencies, however, it is usually possible to set up the required two-dimensional source instead of setting up an array which is equivalent to the source. This is accomplished through the use of a reflector, as described next.

Parabolic Reflector

To understand how a two-dimensional source can be obtained through the use of a reflector, consider the curved line source shown in Fig. 7-15 and examine the radiation pattern, in the plane of the paper, to the right of the source. In the simpler line sources described in the preceding section it was assumed that all parts of the line source operated in phase with each other. This cannot be assumed for the curved line source shown in Fig. 7-15. The phase of the elementary source at *A* may be taken as reference phase. The radiation from any other point on the curved line source is then assumed to be earlier in phase than the radiation from *A*, (the phase difference, in cycles, being equal to the distance in wavelengths between the point in question and straight line AB). Thus, for example, radiation from *B* is in phase with radiation from *A*, radiation from *P* may be one-half cycle ahead of radiation from *A* if distance PQ is one-half wavelength, etc. If the phase of the radiation from each point on the curved line source is chosen correctly, all signals traveling in the direction of arrow *X* will have the same phase as they reach straight line AB. The curved line source is then pre-

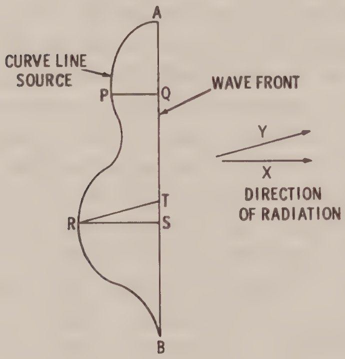


Fig. 7-15. Curved line source.

cisely equivalent to a normal line source, lying along straight line AB, for radiation in direction X.

Now, consider what happens to radiation in the direction of arrow Y. If the angle between the direction of arrow X and the direction of arrow Y is small, distance RT will be nearly the same as distance RS. The phase of radiation crossing straight line AB while traveling in the direction of arrow Y will then be nearly the same as the phase of radiation crossing the line while traveling in the direction of arrow X. In other words, signals will have nearly the same phase at all points on straight line AB, and the curved line source will behave *very nearly* like a normal line source lying along line AB. A system such as the one shown in Fig. 7-15 can be constructed with the help of a reflector, as shown in Fig. 7-16.

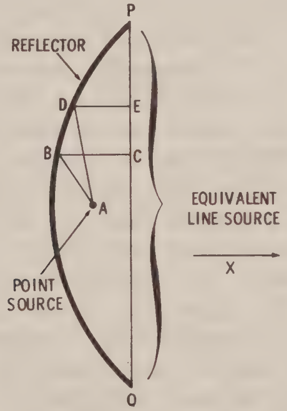


Fig. 7-16. Reflector and equivalent line source.

The desired result will be reached if all signals traveling in the direction of arrow X cross straight line PQ with the same phase. To achieve this the reflector must be shaped so that the sum of distances AB and BC is equal to the sum of distances AD and DE, where B and D are any two points on the reflector. If this is the case, all signals traveling in the direction of arrow X will arrive at the equivalent line source with the same phase (for all signals travel the same distance in reaching the equivalent line source).

A curve that may be used as a shape for the reflector is one called a *parabola* (Fig. 7-17). A parabola may be defined by any straight line, which is called the *directrix* of the parabola, and the point (that is not on the directrix) called the *focus* of the parabola. The parabola itself is a curve that passes through all points the distance of which from the di-

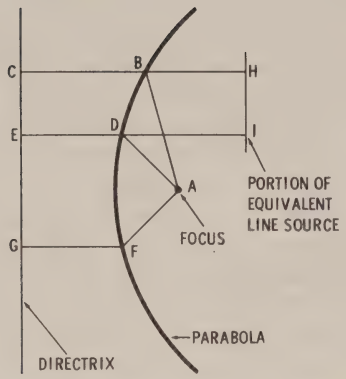


Fig. 7-17. Parabola.

rectrix is equal to their distance from the focus. The distance must be measured along a path perpendicular to the directrix. Thus, *B* lies on the parabola—or the parabola passes through *B*—because distances BA and BC are equal. Similarly *D* lies on the parabola because DA and DE are equal, and *F* lies on the parabola because FA is equal to FG. The usefulness of a parabolic reflector with a point source at the focus of the parabola may be verified by considering line HI, which is parallel to the directrix. Distances CH and EI are equal, because the equivalent line source (the line HI) is parallel to the directrix. Then, since BC is equal to BA, AB plus BH is equal to CH. Similarly, AD plus DI is equal to EI. Since CH and EI are equal, it follows that sum AB plus BH is equal to sum AD plus DI, which was the condition required for a useful reflector (Fig. 7-16).

Paraboloid of Revolution

The antenna system of a tracking radar, or the antenna for a point-to-point communications system, is one that produces a beam like a searchlight. A radiation pattern of this sort is easily obtained from a circular two-dimensional source; the usual antenna consists of a reflector which is equivalent to such a source. The required reflector is a surface that is called, mathematically, a *paraboloid of revolution.* Every parabola has an *axis,* which is a line passing through the focus and perpendicular to the directrix (Fig. 7-18). Imagine now that a parabola is rotated about its axis. It will sweep out a curved surface called a parabola of revolution. If the idea of sweeping out a surface is difficult to grasp, consider a simple

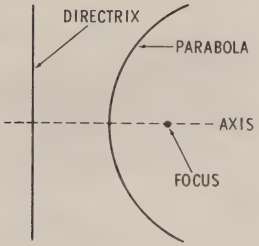


Fig. 7-18. Parabola of revolution.

case. If a circle is rotated about its diameter as an axis, it will sweep out a sphere. The paraboloid reflector is fed from a point source located at the focus, and the resulting system behaves like the desired circular two-dimensional source. It is possible to use a very small section of a very large parabola as a basis for the reflector, as shown in Fig. 7-19A, or a larger section of a smaller parabola as shown in Fig. 7-19B. Reflectors based on these two parabolas will be precisely equivalent. The only difference is in the required point source which feeds energy to the reflector. The feed system, located at the focus, is far in front of the parabola shown in A of Fig. 7-19, and this makes a rather cumbersome mechanical structure. The larger section of a smaller parabola (Fig. 7-19B) makes a more compact assembly but requires a feed system that distributes energy more widely in order that all parts of the *reflector may receive energy. The paraboloid of revolution* usually is called a *dish*, and may be anything from 4 inches to 50 feet in diameter. The reflector can be made lighter by cutting holes in it, leaving only a sort of mesh-work. So long as the holes are small compared to the operating wavelength, this will have relatively little effect on its performance.

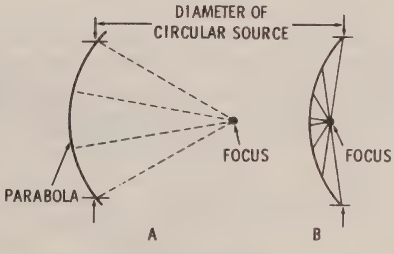


Fig. 7-19. Large and small parabolas of equivalent size.

Parabolic Cylinder

It was pointed out that a circle when rotated about its diameter as an axis sweeps out a spherical surface. In the

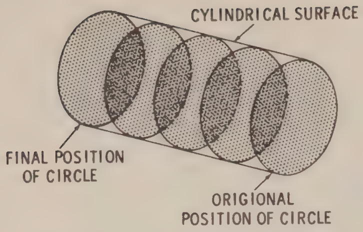

CYLINDRICAL SURFACE
FINAL POSITION OF CIRCLE
ORIGIONAL POSITION OF CIRCLE

Fig. 7-20. Cylinder swept out by a circle.

same way, a parabola rotated about its axis, sweeps out a paraboloid of revolution. If a circle, instead of being rotated, is moved in a direction *perpendicular* to the plane it occupies, it will sweep out a cylindrical surface (Fig. 7-20). If, in the same way, a parabola is moved in a direction perpendicular to the plane in which it lies, it will sweep out a *parabolic cylinder* (Fig. 7-21). In addition, the focus of the parabola

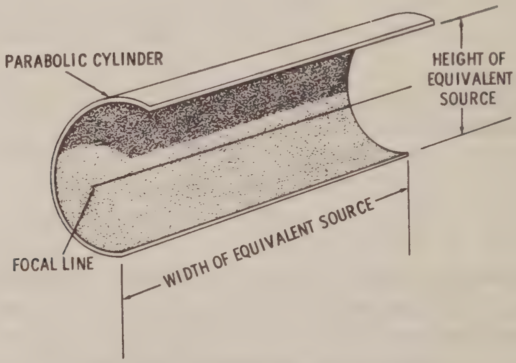

PARABOLIC CYLINDER
HEIGHT OF EQUIVALENT SOURCE
FOCAL LINE
WIDTH OF EQUIVALENT SOURCE

Fig. 7-21. Parabolic cylinder antenna.

will sweep out a *focal line*. A reflector in the shape of a parabolic cylinder may be fed from a line source that is placed along the focal line. The resulting system behaves like a rectangular two-dimensional source. Parabolic cylinders are frequently used in the antenna systems of land-based search radars—the cylinder may be as long as 20 feet. The horizontal beamwidth, which depends on the length of the parabolic cylinder, is usually less than 10°. The vertical beamwidth, which depends on the heighth of the cylinder, is customarily from 3 to 6°. At first glance, the vertical beamwidth appears to be to small for good coverage. The antenna is tipped backward, however, giving the vertical polar diagram shown in Fig. 7-22. On a ship it is necessary to have much greater vertical beamwidth because of the motion of the ship; a nar-

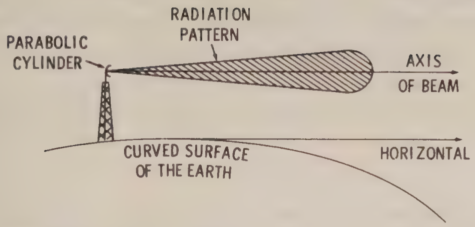

RADIATION PATTERN
PARABOLIC CYLINDER
AXIS OF BEAM
CURVED SURFACE OF THE EARTH
HORIZONTAL

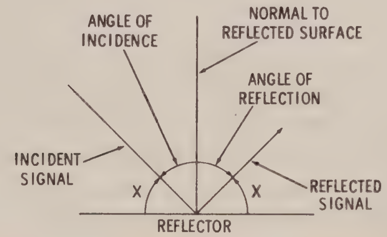
Fig. 7-22. Vertical polar diagram of land-based search radar.

row beam might pass above or below the target. For this purpose, the parabolic cylinder is turned on its end and is made very short. An antenna of this sort is called an *orange peel antenna.*

Optical Design Method

It is possible, instead of using the array design method, to design reflector-type antenna systems by following optical principles. The basis of the optical design method is the law of reflection illustrated in Fig. 7-23. When light strikes a reflecting surface at some angle of incidence, it is reflected in a direction that makes the angle of reflection equal to the angle of incidence. Since the behavior of radio waves is similar to that of light waves, particularly when the reflecting surface is large compared to the wavelength of the signals, it is possible to design a reflector by applying this law. Since the angles of incidence and reflection are equal, and since the normal to the reflecting surface is (by definition) perpendicular to the reflecting surface, the two angles must be equal. It is then possible to design a reflector by the method illustrated in Fig. 7-24.

Fig. 7-23. Law of reflection.


ANGLE OF INCIDENCE
NORMAL TO REFLECTED SURFACE
ANGLE OF REFLECTION
INCIDENT SIGNAL
X
X
REFLECTED SIGNAL
REFLECTOR

If a reflector is illuminated by such a source, energy falling on *A* in Fig. 7-24 will be reflected in the desired direction if the two angles marked X are equal. Similarly, energy falling on B will be reflected in the proper direction if the two angles marked Y are equal, and energy falling on C will be reflected properly if the two angles marked Z are equal. The detailed design of the reflector is shown in Fig. 7-25. Here A is a point on the reflecting surface. Arrow AD points in the direction of the desired radiation pattern beam axis, and arrow AS points toward the source. Then AN is the required normal to the reflecting surface which makes the angle of reflection equal to the angle of incidence, and finally, the direction of the reflecting surface at A must be perpendicular to the normal AN. As might be expected, a reflecting surface that is designed according to this method turns out to be the same parabola described earlier in this section, and the focus of the parabola is, of course, at the position occupied by the source that feeds the reflector.

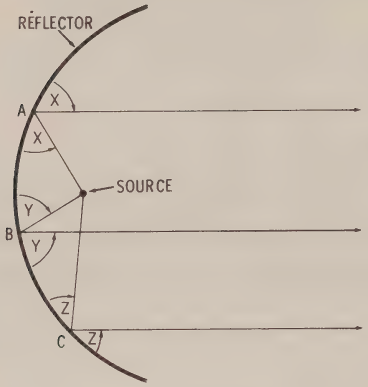


Fig. 7-24. Reflector optical design.

SPECIAL REFLECTORS

The most interesting special reflector is one called a cosecant-squared reflector. This produces a radiation pattern that

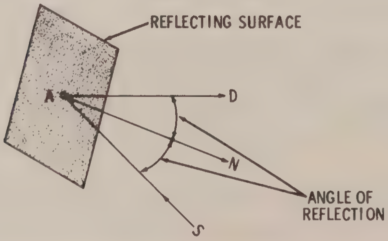


Fig. 7-25. Reflector design details.

is particularly suited to airborne search operations and, to a lesser extent, short-range land-based search radars—so called late-warning, as opposed to early-warning radars. In Fig. 7-26 an aircraft is carrying out an over water search. It is desirable for the radar to give the same response for ships of equal size.

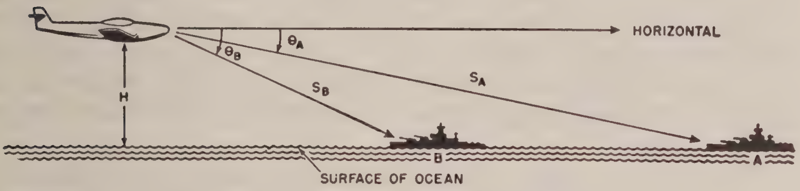


Fig. 7-26. Short-range search problems.

In particular, similar ships at *A* and *B* should produce the same echo response at the output of the radar receiver. The curvature of the earth is neglected in the illustration, since the operating range is assumed to be relatively small—50 miles or less.

Now consider the right-angled triangles formed by the aircraft, the targets, and the point directly beneath the aircraft. An examination of the geometry of Fig. 7-26 will show that *H*, the altitude of the plane, equals S_A the distance to the farther ship times the sine of angle Θ_A. Similarly expressed, the equation is $H = S_A \text{ Sin } (\Theta_A)$. Similarly, H also equals $S_B \text{ Sin } (\Theta_B)$, where S_B is the distance to the near ship, and Θ_B is the angle. By division, the two equations can be written also as follows:

$$S_A = \frac{H}{\text{Sin } (\Theta_A)}$$

and,

$$S_B = \frac{H}{\text{Sin } (\Theta_A{}^{\circ})}$$

Now, the reciprocal of the sine of an angle is the cosecant of the angle expressed, $C_{sc}(\Theta) = 1/\text{Sin}(\Theta)$. By substitution, the equations now can be written as follows

$$S_A = (H)\ C_{sc}\ (\Theta_A)$$

and,

$$S_B = (H)\ C_{sc}\ (\Theta_B),$$

and reduced to a general equation,

$$S = (H)\ C_{sc}\ (\Theta),$$

where S is the distance (slant range) from plane to ship (target), H is the altitude of the plane and Θ is the angle (depression angle) between the surface of the sea and the radio line of sight to the ship. That is, the slant range to a target is proportional to the cosecant of the depression angle to the target.

It has been pointed out that a one-way signal level plot for a radar is also a plot of relative range, for identical targets, on targets in different directions for a given receiver output. The desired antenna system for operations of this sort is, therefore, one in which the signal level in directions below the horizontal is proportional to the cosecant of the angle below the horizontal. The power level is proportional to the square of the cosecant of the angle, and the antenna derives its name from this fact. (Most of the discussion is in terms of power, although almost all practical work is in terms of signal level.) The desired radiation pattern (one-way signal level plot) is shown in A of Fig. 7-27, and a frequently used pattern is shown in B. The pattern is obtained by adding an auxiliary reflector which sends some of the energy from the feed system at an angle a little below the horizontal.

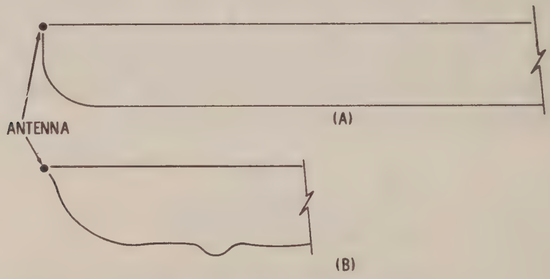
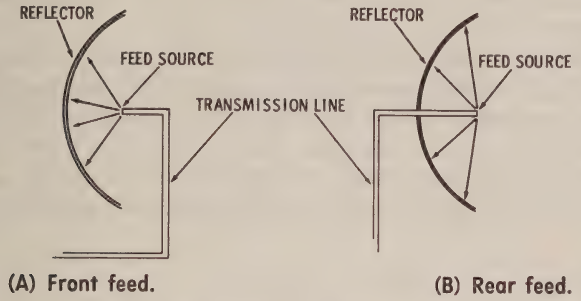


Fig. 7-27. Cosecant Patterns.

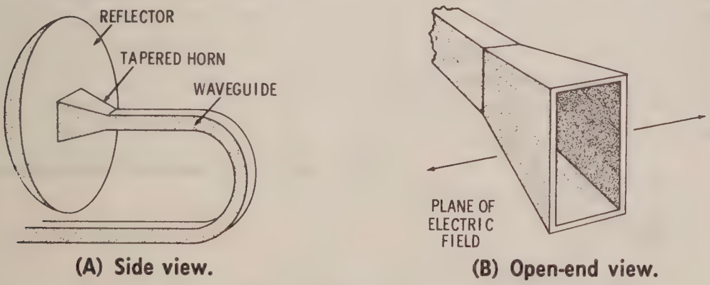
REFLECTOR FEED SYSTEMS

Feed systems for reflector-type antenna systems may be divided into front feeds and rear feeds. In a front-feed system, the transmission line (rigid coaxial line or waveguide) is led around the edge of the antenna to the front, where it is terminated by a radiator of some sort, as shown in Fig. 7-28A. In a rear feed, the transmission line is passed through the reflector from the rear, as in Fig. 7-28B. The choice between front and rear feed systems is usually made on the basis of mechanical difficulties. The ordinary front-feed system is nothing more than a tapered horn several wavelengths long. The taper serves two purposes: it acts as a tapered impedance-matching section such as that described previously and minimizes the mismatch between the impedance of space and that of the waveguide. It also allows the small waveguide, the dimensions of which are fixed by the necessity of avoiding operation in undesired modes, to feed a rather large two-dimensional source. Fig. 7-29A shows a tapered horn feeding a paraboloid of revolution, and Fig. 7-29B shows the horn itself. The system shown in the illustration is arranged to radiate horizontally polarized signals, as may be seen from the fact that the narrow dimension of the waveguide is horizontal. The open end of the horn, which acts as a two-



(A) Front feed. (B) Rear feed.

Fig. 7-28. Front and rear feed systems.



(A) Side view. (B) Open-end view.

Fig. 7-29. Tapered horn.

dimensional source, is located at the focus of the paraboloid.

The dimensions of the open end of the horn are chosen to obtain the desired radiation pattern, which will distribute energy more or less uniformly over the reflector with a minimum of loss. Occasionally, to obtain more uniform illumination of the reflector, a horn of circular cross section is used. This is ordinarily fed by a circular waveguide operated in the $TE_{1,1}$ mode (which is equivalent to the $TE_{0,1}$ mode in a rectangular guide).

A *septum*, or conducting partition, is inserted in the circular waveguide in order to control the plane of polarization. This operates by absorbing all signals having electric fields parallel to the surface of the septum. The horn shown in Fig. 7-30, for example, radiates only horizontally polarized signals. A *sectoral horn* is used to feed an orange peel antenna.

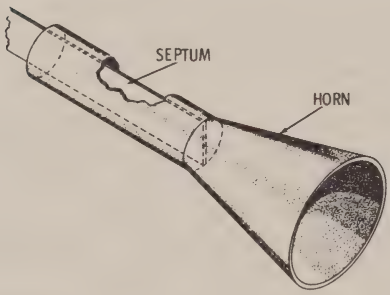


Fig. 7-30. Circular horn fed by circular guide with septum.

Rear Feeds

The usual rear feed is a dipole that is fed by radiation from a waveguide, and a reflector, as shown in Fig. 7-31. The waveguide is slightly tapered in order to obtain an approximate impedance match to the dipole which it feeds. The dipole, which is simply a half-wave antenna, is supported on the septum that extends from the end of the waveguide. The

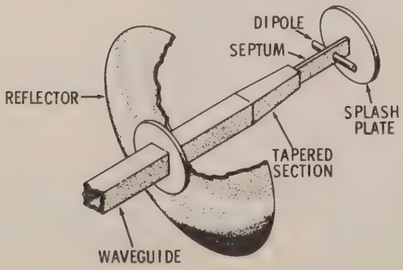
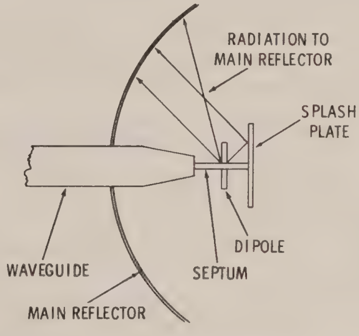


Fig. 7-31. Rear feed with splash plate.

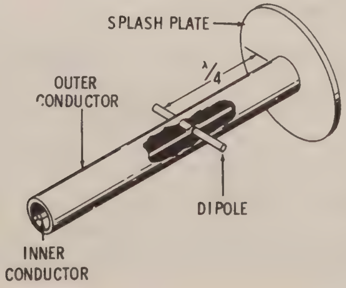
small reflector, called a *splash plate,* reflects energy from the dipole back toward the main reflector. A simplified illustration of the same arrangement is shown in Fig. 7-32. The splash plate sometimes is replaced by another radiating element (dipole). The radiation from the waveguide sets up electric fields in the neighborhood of the two dipoles. These fields induce currents in the dipoles, and the currents cause the dipoles to radiate. By adjustment of the length of a dipole of this sort, it is possible to change its impedance and, consequently, the phase of its current and radiation. By suitable choice of lengths and spacing, the radiation from the two dipoles can be made to fall on the main reflector. Various other combinations are also used. More than two dipoles sometimes are seen, and, occasionally, a combination of two or more dipoles may be used with a splash plate. All of these, however, serve the same purpose; causing the radiation from the waveguide to fall uniformly on the main reflector.

Fig. 7-32. Schematic of rear-feed system.


RADIATION TO MAIN REFLECTOR
SPLASH PLATE
DIPOLE
SEPTUM
WAVEGUIDE
MAIN REFLECTOR

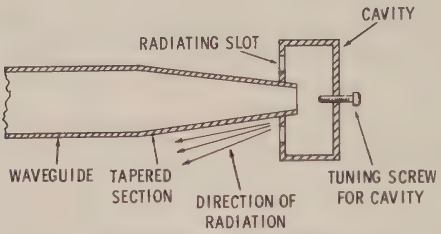
The method of feeding a dipole and splash plate from rigid coaxial line is shown in Fig. 7-33. One side of the dipole is connected to the outer conductor; the other side is connected to the inner conductor of the coaxial line and protrudes

Fig. 7-33. Coaxial feed to dipole.

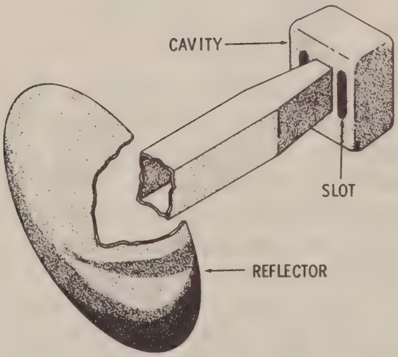

SPLASH PLATE
$\lambda/4$
OUTER CONDUCTOR
DIPOLE
INNER CONDUCTOR

through an opening in the outer conductor. The splash plate serves the double purpose of reflector and short-circuit termination from the coaxial line (1 quarter-wavelength beyond the dipole).

Another useful rear feed is the *Cutler* feed shown in Fig. 7-34. This consists of a resonant cavity, which is tuned by means of the adjusting screw, with two radiating slots which feed energy to the reflector. The radiation elements of any of the rear feeds described here may be inclosed in an airtight box of polystyrene for the purpose of pressuring the waveguide or coaxial line.



(A) Cross-sectional view.



(B) Perspective view.

Fig. 7-34. Cutler feed.

Linear Arrays

A linear array is required to feed a parabolic cylinder reflector of the type shown in Fig. 7-21. The usual array consists of a set of probe-fed dipoles. Each dipole is, in effect, a center-fed half-wave antenna (Fig. 7-35). One half of each dipole is fed by the probe, which projects into the waveguide parallel to the electric field. The other half of the dipole is fed by transfer of energy across the quarter-wave two-wire transmission line which supports the two halves of the dipole.

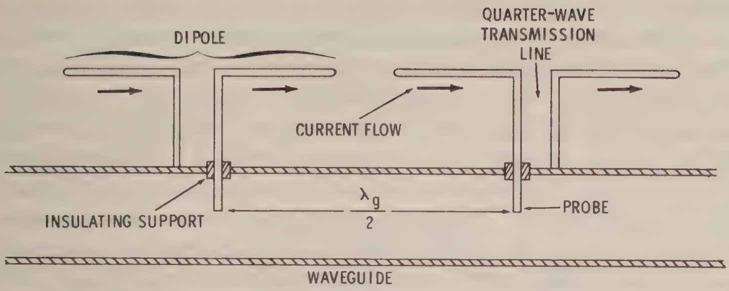

QUARTER-WAVE TRANSMISSION LINE
DIPOLE
CURRENT FLOW
INSULATING SUPPORT
$\frac{\lambda_g}{2}$
PROBE
WAVEGUIDE

Fig. 7-35. Probe-fed dipoles.

Successive pairs of dipoles are spaced at intervals of one-half guide-wavelength along the waveguide. Note that adjacent dipoles are oriented in opposite directions so that they will radiate in phase. This result is obtained because the excitation of adjacent dipoles coming from parts of the waveguide one-half guide-wavelength apart, is precisely out of phase. Reversing the direction of the probe-fed and passive halves of the adjacent dipoles, therefore, results in current flow in the same direction, at any instant, in all radiating elements. As the input end of the linear array shown in Fig. 7-36, the energy level in the waveguide is relatively high and the probes extend only a short distance into the waveguide. Near the other end of the array the energy level is lower because energy has been withdrawn by the earlier dipole probes and the probes extend farther into the waveguide. To avoid any reflection, the waveguide is terminated in its characteristic impedance by a dissipative termination.

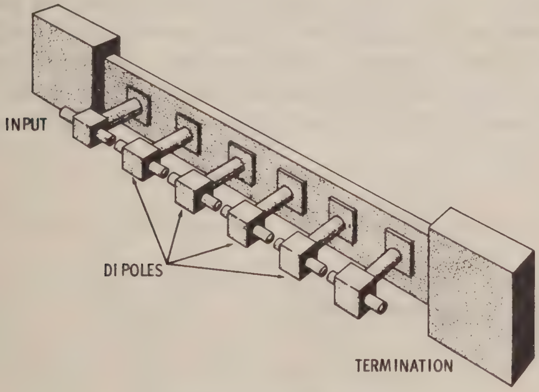
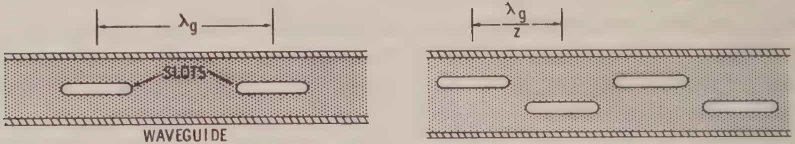

INPUT
DIPOLES
TERMINATION

Fig. 7-36. Linear array.

Another type of linear array, which is more rugged but more difficult to design, is made by simply cutting slots in a waveguide. The slots radiate, exactly like the slots in a resonant cavity, and allow energy to leak out of the waveguide. A linear array of this sort, called a *leaky waveguide*, is shown in Fig. 7-37A, where the slots are spaced 1 guide-wavelength apart in order that the radiation from adjacent slots may be of the same phase. The waveguide, of course, is operated in the $TE_{0,1}$ mode. A better radiation pattern is obtained by spacing the slots only one-half guide-wavelength as in Fig. 7-37B. Here, the waveguide is operated in the $TE_{0,2}$ mode and adjacent slots are cut in alternate sides of the guide so that, as before, radiation from all slots is in the same phase.

The complexity of the system lies in the difficulty of controlling the amount of radiation from a given slot. In practice, the slots are not as simple as those shown in the illustration. As with the probe-fed dipole array, the waveguide is terminated in a dissipative termination, the impedance of which matches the characteristic impedance of the wave guide. A somewhat similar array is sometimes used for radar beacon antennas. A beacon antenna frequently is required to be omnidirectional, and also must send and receive horizontally polarized signals. One method of obtaining such a pattern is to use a circular waveguide operating in the $TM_{0,1}$ mode and feed a number of dipoles. Fig. 7-38A shows the



(A) Slots spaced 1 guide wavelength apart.

(B) Slots spaced ½ guide wavelength apart.

Fig. 7-37. Leaky waveguide.

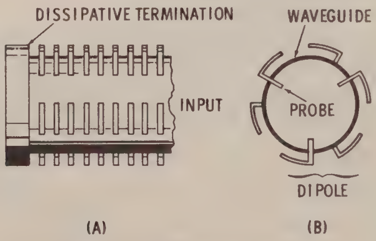


Fig. 7-38. Antenna array.

general configuration of the antenna array; Fig. 7-38B shows one set of dipoles and the probes which feed them. As with the probe-fed dipole array described above, the adjacent sets of dipoles are spaced one-half guide-wavelength along the waveguide, and the orientation of the dipoles is reversed in successive sets.

METAL LENSES

Most microwave antennas consist of reflectors with suitable feed systems, as described in the earlier portions of this chapter. There are, however, a number of types of microwave antennas which operate without reflectors. The simplest of these is the tapered horn antenna which terminates a waveguide transmission line, and is similar to the tapered horn used in front feed systems for a paraboloid revolution. This if often used with microwave test equipment. Another nonreflector antenna is the radar beacon antenna described above. Still another type of antenna, the *metal lens* antenna, frequently is used in microwave communications installations. This has no reflector, but makes use of a metal lens to focus the radiated energy into a highly directive beam. The same system, of course, provides highly directive reception. The basic theory of the metal lens antenna system is illustrated in Fig. 7-39. If a waveguide is terminated by the usual tapered horn, the radiation from the horn will be compressed moderately into a beam. The beam is, however, divergent, or spreading, and the wave fronts in the beam, which are everywhere perpendicular to the direction in which the signal travels, will be curved as shown in the figure. Suppose, now,

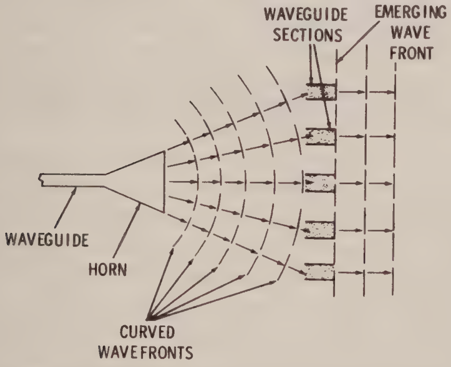
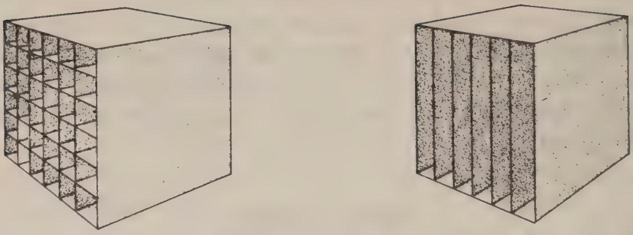


Fig. 7-39. Schematic of the basic theory of a metal lens.



(A) Waveguide sections beside one another.

(B) Tops and bottoms of waveguide sections omitted.

Fig. 7-40. Metal lens developed from waveguide sections.

that this radiation falls on a number of waveguide sections that are not all of the same length. As pointed out in Chapter 3, the phase velocity of a signal in a wavelength is greater than the phase velocity of the signal in space. (In unobstructed space, both phase and group velocity are the same.) Signals that travel through long sections of waveguide are advanced in phase in relation to those that travel through short sections. Since the wavefront is a surface of constant phase and since signals have the same phase at every point on a wave front, this passage through waveguide sections of different lengths alters the shape of the wavefronts. By proper design, the wavefront of the emerging signals can be made flat and, therefore, identical with the wave front of the signal from a two-dimensional source. Although only a few waveguide sections are shown in Fig. 7-39, it is assumed that the sections are placed beside one another so that they appear as shown in Fig. 7-40A, when viewed from the horn. The phase velocity through a waveguide, however, depends only on the

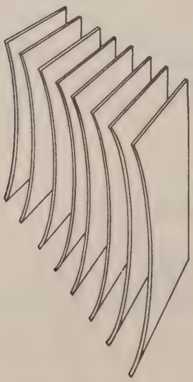


Fig. 7-41. Square metal lens.

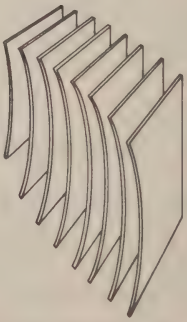


Fig. 7-42. Circular metal lens.

width of the waveguide (for propagation in the $TE_{0,1}$ mode), and the tops and bottoms of the waveguide sections may be omitted as shown in Fig. 7-40B without altering the behavior of the system.

A square metal lens is shown in Fig. 7-41 and a round metal lens is shown in Fig. 7-42. There are, of course, many more sections than are shown in the illustration. The feedhorn may be small, as in Fig. 7-39, or it may be extended until it reaches the lens.

Merit of Metal Lens Systems

The merit of a metal lens antenna is obvious. Since energy is fed *through* the lens, instead of being reflected from it, the feed horn may be extended as far as is necessary to guarantee that no energy is lost around the edges of the lens. This cannot be done with a reflector system, for the large feed system would interfere with the signals *from* the reflector, or, in a receiving antenna, with the signals *to* the reflector. Another advantage of the metal lens is its simplicity of construction. Instead of a precisely curved reflector, only a number of flat metal strips are used. The chief disadvantage of the metal lens system is its weight. Since, by the time the lens and the associated supporting structure is completed it is likely to weigh more than the equivalent reflector, the use of metal lenses is confined largely to fixed stationary antennas of the type used in communications installations.

SCANNING AND STABILIZATION

Almost all radar installations are required to *scan*—that is, to move their antennas in such a way that the beam sweeps through a certain predetermined space. A search radar scans in order to look for targets, and a tracking radar scans in order to find out the precise direction of the target under observation. The basic scanning method used by search radars and the most usual one, is called PPI (plan-position-indicator) scan. The term was coined by the British in the early days of radar.

PPI Display

A PPI display uses a brilliance-modulated cathode-ray tube, one in which the spot on the screen normally is extinguished (by cutting off the electron beam). Information is displayed by allowing the receiver output to turn the spot on. A double layer phosphor is used. The electron beam, when it is turned on, produces a bright blue flash in one layer of the phosphor, and this flash produces an orange glow in the second layer. The blue flash disappears almost immediately when the electron beam is turned off, but the orange glow persists for some time. In addition, the orange glow is not brought to full brilliance until there have been a number of blue flashes in a short time interval, at nearly the same point. The echo from a target appears each time at nearly the same point on the cathode-ray tube screen, whereas noise may appear at any point and is unlikely to appear very often, during a short time interval, at any given point. The orange glow from a target is therefore brighter than the glow from the radar-receiver noise output. The screen is observed through an orange filter so that the blue flashes are not seen.

The spot starts at the center of the screen after each transmitter pulse and moves radially outward. The position of the bright spot produced by an echo is, then, determined by the range to the target which gives rise to the echo. Nearby targets produce spots near the center of the screen and distant targets produce spots near the edge. The direction in which the spot moves corresponds to the direction in which the antenna is looking. The direction and range of a target then are determined immediately from the position of the spot on the screen, as shown in Fig. 7-43. In PPI scanning, the antenna rotates continuously.

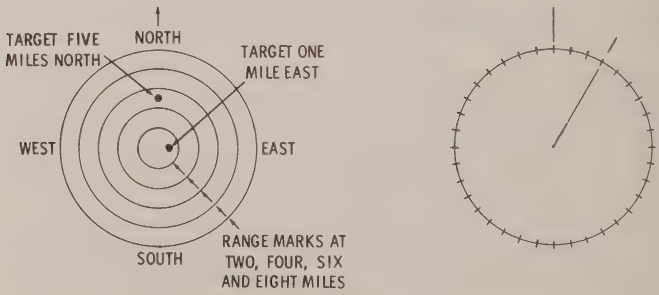


Fig. 7-43. PPI display.

Azimuth Stabilization

For a land-based search radar, the display can be oriented in any convenient direction. In general, the region of greatest interest is shown at the top of the screen, or at the point

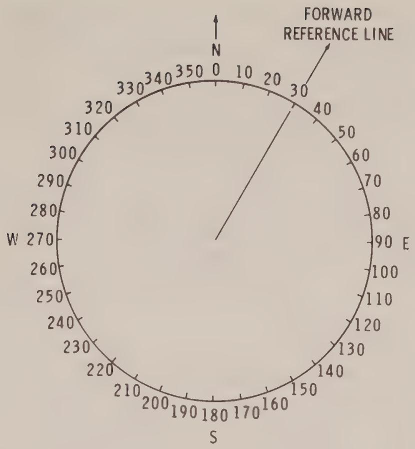


Fig. 7-44. Azimuth-stabilized PPI.

farthest from the observer, marked north in Fig. 7-43. In mobile equipment, the upper portion of the screen usually shows targets ahead of the ship or aircraft which carries the radar. The right-hand side of the screen then shows targets to *starboard* (to the right when facing forward) and the left-hand side shows targets to *port* (to the left when facing forward). The lower portion, of course, shows targets behind the ship or aircraft. For many purposes, however, it is desirable to show North at the top of the screen so that the picture does not change when the ship or aircraft is turned. This can be done by making the PPI deflection system track ahead of the antenna or behind it, as required. It is customary to show a bright line on the screen in the forward direction (Fig. 7-44). It is also usual to show directions by means of numbers, in degrees, instead of by North, East, South, and West. In some cases, forward is shown at the top of the screen and a reference line is added to show north. This, however, is not a stabilized scan, since the picture will rotate when the ship or aircraft turns. A stabilized scan such as this is known as an *azimuth-stabilized scan.*

Attitude Stabilization

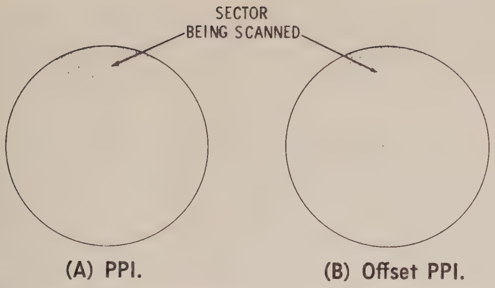
Consider a situation where an aircraft is observing a target which is directly ahead and at a considerable angle of depression. Suppose that the aircraft is pointing properly toward the target so that its radar receives an echo when the radar looks directly ahead. If the aircraft drops its right wing in making a turn, the center axis of the radar beam no longer will fall on the target when the radar points ahead, but will fall on some other point on the ground. The target then will appear to be on the right instead of directly forward. The resulting error, which is called an *attitude error* may be corrected by suitably modifying the motion of the antenna and the PPI deflection system. A scanning system in which this error is corrected is said to be attitude stabilized. The equations for attitude stabilization are rather involved and need not be detailed here.

Sector Scan

In sector scanning, the antenna sweeps back and forth over a sector of a circle (Fig. 7-45) instead of around the complete circle. A heavy antenna must be driven back and forth, but a light antenna sometimes uses a *resonant-scanning* mount in which it oscillates back and forth under the influence of springs. The mechanical system (antenna and springs) can be made to resonate at the scanning frequency so that the drive motor, to keep the antenna in motion, is only required to overcome the frictional forces. The vertical position of a sector-scanning antenna is sometimes changed during the scanning cycle. Motion of this sort is called *nodding*. In Fig. 7-45, the antenna looks higher during sweeps to the right than during sweeps to the left, and thus is enabled to cover a larger region.

Fig. 7-45. Sector of a circle.

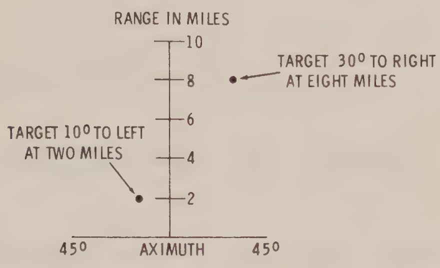
The output of a sector-scanning radar may be displayed on an ordinary PPI, as in Fig. 7-46A, or on an offset PPI in which the center of the pattern is displaced from the center of the screen, as shown in Fig. 7-46B. The center of the pattern is sometimes shifted completely off the screen, and the display, then, is called a PEPPI (partial expanded PPI). The usual display for a sector scan, however, is a B-scope. Here the spot moves vertically during the interval after each pulse, but the whole trace is displaced from side to side as the antenna scans. The result, shown in Fig. 7-47, is a plot in which


SECTOR
BEING SCANNED

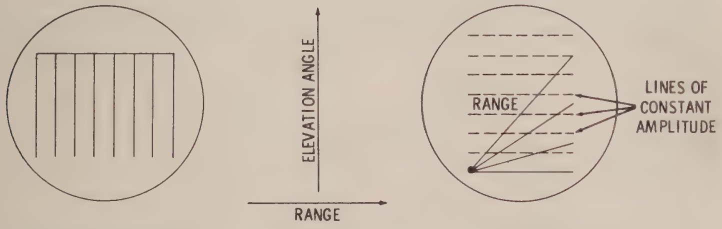
(A) PPI. (B) Offset PPI.

Fig. 7-46. PPI and offset PPI for sector being scanned.

Fig. 7-47. B-scope plot.


RANGE IN MILES
10
8
6
4
2
TARGET 30° TO RIGHT
AT EIGHT MILES
TARGET 10° TO LEFT
AT TWO MILES
45°
AXIMUTH
45°

the vertical coordinate is range and the horizontal coordinate is azimuth (angle). Attitude stabilization occasionally is used with sector scanning, but azimuth stabilization is rarely seen. Occasionally, an antenna scans through a sector in a vertical plane. This motion is called vertical-sector scanning, and it allows the radar to determine the range and elevation angle


ELEVATION ANGLE
RANGE
RANGE
LINES OF
CONSTANT
AMPLITUDE

(A) B-scope equivalent. (B) Offset PPI.

Fig. 7-48. Vertical sector scan.

(angle above the horizontal) of a target. The radar output may be displayed on the equivalent of a B-scope as in Fig. 7-48A, or on the equivalent of an offset PPI as in Fig. 7-48B.

Appendix I

MATHEMATICAL CONSTRUCTION OF PROFILE DIAGRAM

Height of the Earth's Surface Above Straight Line Joining Two Points

1. In order to construct a profile diagram it is necessary to know the maximum height of the earth's surface above a straight line between two points on the surface. In A of Fig. I-1, *P* and *Q* are two point on the surface of the earth and *S* is the distance between them. The straight line joining *P* and *Q* is a straight line through the earth, and the arc joining these points represents the surface of the earth. *S* may be measured either along the arc joining *P* and *Q* or along the chord (the straight line), since the two measurements are nearly equal, differing by less than

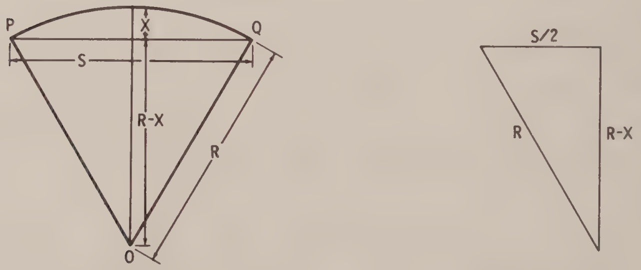


Fig. I-1. Mathematical construction of profile diagram.

1% when *S* is 400 miles and by even less when *S* is smaller. *R* is the radius of the earth, and *X* is the maximum distance between the arc and the chord. Point *O* is the center of the earth.

2. Half of the diagram is shown in B of Fig. I-1 as a right triangle with a hypotenuse of length *R* and legs of length (S ÷ 2) and R – X). Then, by Pythagoras' theorem, which states that the sum of the squares of the legs of a right triangle is equal to the square of the hypotenuse:

$$\frac{S^2}{4} + (R-X)^2 = R^2$$

(1st leg) (2nd leg) (hypotenuse)

$$\frac{S^2}{4} + R^2 - 2RX + X^2 = R^2$$

$$\frac{S^2}{4} = 2RX - X^2$$

$$\frac{S^2}{4} = X\,(2R - X)$$

The second factor on the right, (2R – X), may be replaced by (2R). This can be accomplished without introducing any appreciable error, because *X* is negligible compared with 2R. When this substitution has been carried out, the equation becomes

$$\frac{S^2}{4} = 2RX$$

$$X = \frac{S^2}{8R}$$

3. Both *S* and *R* are in miles, and so, therefore, is *X*. Since *X* is desired in feet, the right side of the equation must be multiplied by 5,280. The equation then becomes:

$$Xf = \left(\frac{S^2}{R}\right)\left(\frac{5280}{8}\right) = \left(\frac{S^2}{R}\right)(660)$$

where,
X_t is the distance in feet. If *R* is set equal to 4,000 miles, the normal radius of the earth, the equation becomes:

$$X_t = S^2\left(\frac{660}{R}\right) = S^2\left(\frac{660}{4000}\right) = S^2\left(\frac{33}{200}\right)$$

$$X_t = S^2\left(\frac{99}{600}\right)$$

But 99/600 is nearly equal to 100/600 or 1/6, so that an approximation that is correct for all practical purposes may be expressed by

$$X_t = \frac{S^2}{6}$$

The effect of substitution 4/3R, the radius of the modified earth, for *R* the radius of the normal earth, is to multiply the denominator of the final expression by the factor 4/3, and 4/3 times 6 is equal to 8. Consequently,

$$X_t = \frac{S^2}{8}$$

for the modified earth.

Appendix II

WAVEGUIDE MATHEMATICS AND EQUIVALENT CIRCUITS OF AMPLIFIER AND QUARTZ CRYSTAL

Mathematics Involved in Theory of Stub Tuning

The original equation is:

$$Z = \frac{E}{I} = \frac{E_1 + E_2}{I_1 - I_2}$$

$$(E_1 + E_2) = Z(I_1 - I_2)$$

but,

$$I_1 = \frac{E_1}{Z_o} \text{ and } I_2 = \frac{E_2}{Z_o}$$

so that,

$$(E_1 + E_2) = Z\left(\frac{E_1}{Z_o} - \frac{E_2}{Z_o}\right)$$

$$= \frac{Z(E_1 - E_2)}{Z_o}$$

and,

$$Z_o(E_1 + E_2) = Z(E_1 - E_2)$$

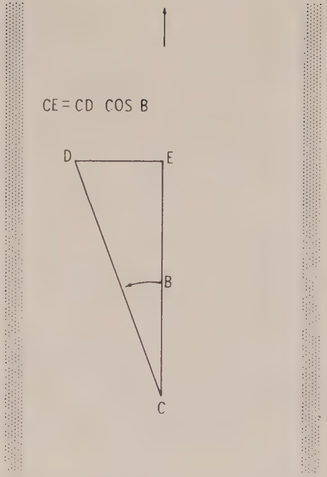


Fig. II-1. Geometry of equation for derivation of group velocity.

$$(Z + Z_o)E_2 = (Z - Z_o)E_1$$

and the final step brings out the equation,

$$E_2 = \frac{E_1(Z - Z_o)}{(Z + Z_o)}$$

Group Velocity

The group velocity (Fig. II-1) is given by the equation:

$$v_g = c \cos B$$

Consider a portion of the signal that moves to the left. At some instant, this portion is at *C*. After a short time, *t*, the electromagnetic wave will have moved to *D*, where distance CD is equal to ct. At the same time, the wave travels a distance CE *along* the waveguide, where CE is equal to $v_g t$. Then,

$$CE = v_g t = CD \cos B = ct \cos B,$$

and,

$$v_g t = ct \cos B$$

or,

$$v_g = c \cos B$$

Energy, therefore, travels *along* the waveguide more slowly than it does through unobstructed space, in spite of the fact that the individual electromagnetic waves travel with a velocity, *c*.

Phase Velocity

In a steady state, when the waveguide carries a continuing flow of energy, the pattern moves with the phase velocity given by the equation:

$$v_p = \frac{c}{\cos B}$$

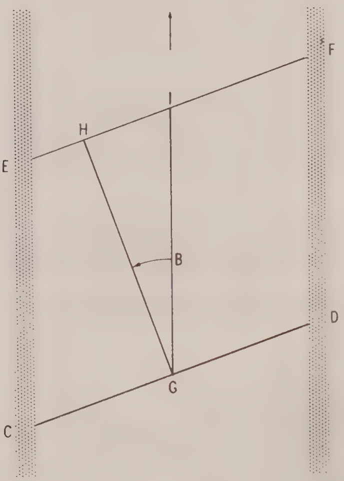


Fig. II-2. Geometry of equation for derivation of phase velocity.

The diagram in Fig. II-2 shows this clearly. Line CD is a wave front of the electromagnetic wave that moves to the left. If G is a point on this wave front at some instant, it will have moved to *H* on wave front EF after a time interval *t*, where distance GH is equal to ct. The corresponding point on the field pattern, however, moves to *I* in the same interval, where distance GI is equal to $v_p t$. Then,

$$GH = ct = GI \cos B = v_p t \cos B$$

whence,

$$ct = v_p t \cos B$$

$$c = v_p \cos B$$

and,

$$v_p = \frac{c}{\cos B}$$

The field pattern, therefore, moves faster than the electromagnetic waves. This, of course, might have been expected from the fact that λ_g is greater than λ_a, since the velocity of any wave system is equal to the product of its frequency by its wavelength, and the frequency is unchanged, so that

$$c = f \lambda_a$$

and,

$$v_p = \frac{C}{\cos B} = \frac{f \lambda_a}{\cos B} = f \lambda_g$$

Characteristics of $TE_{o,n}$ Modes

For any $TE_{o,n}$ mode, angle *B* is given by the equation:

$$\sin B = \frac{n \lambda_a}{2b}$$

Since the cutoff wavelength is the one for which sin B is 1,

$$\lambda_c = \frac{2b}{n}$$

for *any* $TE_{o,n}$ mode. The relation between λ_g and λ_a, as well as the relations between v_g, v_p, and c, depend only on angle *B*. They will, therefore, hold for any $TE_{o,n}$ mode, so that

$$\lambda_g = \frac{\lambda_a}{\cos B}$$

$$v_g = c \cos B$$

$$v_p = \frac{C}{\cos B}$$

The $TE_{m,o}$ modes are produced by two sets of horizontally polarized waves which travel down the guide along paths that are parallel to the sides and that makes angle *A* with the top and bottom. Energy will be carried if the angle satisfies the equation:

$$\sin A = \frac{m\,\lambda_a}{2_a}$$

where a is the height of the waveguide.

The patterns are similar to those of the $TE_{o,n}$ modes except that they are uniform across the guide and vary from top to bottom (and, of course, the electric fields are horizontal because the waves are horizontally polarized). A of Fig. II-3 shows the electric field pattern for the $TE_{1,0}$ mode; B shows the same pattern for the $TE_{2,0}$ mode. The cutoff wavelength is, of course, the wavelength for which sin A is equal to 1, so that λ_c is equal to 2a/m. The other characteristics are developed in the same way as those of a $TE_{o,n}$ mode, except that angle *A* is used in place of angle B. Then,

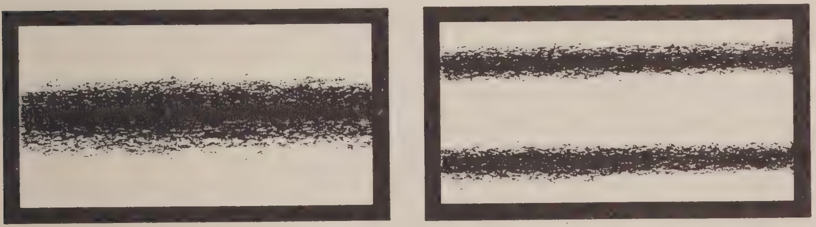


Fig. II-3. Electric field patterns for $TE_{m,o}$ modes.

$$\lambda_g = \frac{\lambda_a}{\cos A}$$

$$v_g = c \cos A$$

$$v_p = \frac{C}{\cos A}$$

Characteristics of Complete Pattern

The two simple patterns combine to form a more complex pattern. The wavelength of the combined pattern is equal to $\lambda/\cos A$, where λ is the wavelength of each of the simple patterns measured along their paths. Since λ is also equal to $\lambda_a/\cos B$, the wavelength of the combined pattern is:

$$\lambda_g = \frac{\lambda_a}{\cos A \ \cos B}$$

The combined pattern moves with a phase velocity equal to $1/\cos A$ times the velocity of the simple patterns, so that:

$$v_p = \frac{c}{\cos A \ \cos B}$$

In the same way, the group velocity of the combined pattern is cos a times the speed with which energy travels along the paths of the simple patterns, and

$$v_g = c \cos A \ \cos B$$

Finally, the cutoff wavelength for the combined pattern is the wavelength (of the signal, not the pattern) for which sin A is equal to 1. Then, $m\lambda/2a$ is equal to $\lambda_a/2a \cos B$ which is equal to 1 when λ_a is equal to the cutoff wavelength λ_g. Then, λ_c is equal to $(2a/m) \cos B$, and this, when defeloped, leads to the expression

$$\lambda_c = \frac{1}{\sqrt{(m/2a)^2 + (n/2b)^2}}$$

This method of building up the combined pattern is reasonable if the four sets of waves are polarized vertically, because the two simple patterns are similar to those of a $TE_{o,n}$ mode that is polarized vertically. If the four sets of waves are polarized horizontally, two simple patterns can be built up similar to those of the $TE_{m,o}$ mode. These will move to the right and left, and can be combined to find the complete pattern. The combined pattern formed by vertically polarized waves will be different from that formed by horizontally polarized waves, but the cutoff and guide wavelengths and the phase and group velocities will be the same. It is worth noting that:

$$\frac{1}{\sqrt{(m/2a)^2 + (n/2b)^2}}$$

is smaller than either of the quantities:

$$\frac{1}{\sqrt{(m/2a)^2}} = \frac{1}{(m/2a)} = \frac{2a}{m}$$

or,

$$\frac{1}{\sqrt{(n/2b)^2}} = \frac{1}{(n/2b)} = \frac{2b}{n}$$

Therefore, if λ_a is smaller than λ_c, it is surely smaller than the cutoff wavelengths for the simple patterns in either the vertically polarized case or the horizontally polarized case.

Cutoff Wavelength

The discussion earlier in this appendix shows that the cutoff wavelength, λ_c, is given by:

$$\frac{1}{\sqrt{(m/2a)^2 + (n/2b)^2}}$$

for either a $TE_{m,n}$ mode or a $TM_{m,n}$ mode. Since the cutoff wavelength for a $TE_{m,o}$ mode is:

$$\frac{2a}{m} = \frac{1}{(m/2a)} = \frac{1}{\sqrt{(m/2a)^2}}$$

and that for a $TE_{o,n}$ mode is:

$$\frac{2b}{n} = \frac{1}{(n/2b)} = \frac{1}{\sqrt{(n/2b)^2}}$$

the expression:

$$\lambda_c = \frac{1}{\sqrt{(m/2a)^2 + (n/2b)^2}}$$

can be used for *any* existing mode. It also shows that the cutoff wavelength for a given waveguide becomes shorter when either *m* or *n* is raised. It is, of course, necessary to remember that *there is no mode for which both m and n are zero,* and that *there are no TM Modes for which either m or n is zero.*

Characteristics of Complete Pattern

The mathematics involved in the expression for the cutoff wavelength is:

$$\lambda_c = \left(\frac{2a}{m}\right) \cos B$$

but,

$$\cos I = \sqrt{1 - \sin^2 B}$$

and,

$$\sin B = \frac{n\,\lambda_a}{2b}$$

At the cutoff wavelength, λ_a is equal to λ_c, so that:

$$\sin B = \frac{n\,\lambda_c}{2b}$$

$$\cos B = \sqrt{1 - \left(\frac{n\,\lambda_c}{2b}\right)^2}$$

and,

$$\lambda_c = \left(\frac{2a}{m}\right) \sqrt{1 - \left(\frac{n\,\lambda_c}{2b}\right)^2}$$

$$\left(\frac{m\,\lambda_c}{2a}\right) = \sqrt{1 - \left(\frac{n\,\lambda_c}{2b}\right)^2}$$

$$\left(\frac{m\,\lambda_c}{2a}\right)^2 = 1 - \left(\frac{n\,\lambda_c}{2b}\right)^2$$

$$\left(\frac{m\,\lambda_c}{2a}\right)^2 + \left(\frac{n\,\lambda_c}{2b}\right)^2 = 1$$

$$(\lambda_c)^2 \left[\left(\frac{m}{2a} \right)^2 + \left(\frac{n}{2b} \right)^2 \right] = 1$$

$$(\lambda_c)^2 = \frac{1}{\left[\left(\frac{m}{2a} \right)^2 + \left(\frac{n}{2b} \right)^2 \right]}$$

$$\lambda_c = \frac{1}{\sqrt{\left(\frac{m}{2a} \right)^2 + \left(\frac{n}{2b} \right)^2}}$$

Equivalent Circuit of Single-Tube Amplifier

The behavior of an ordinary vacuum tube used as an amplifier normally is described by three constants. These are *plate resistance* R_p, *mutual conductance* G_m, and *amplification factor* μ. The last of these is the Greek letter mu. The same letter is used as an abbreviation for the prefix micro (one millionth) as in μamp or μv for microampere or microvolt.

The current passing through a vacuum tube depends on the plate-to-cathode voltage and on the grid-cathode voltage. It is assumed, for simplicity, that the screen-cathode voltage is held fixed. This can be accomplished readily by means of a bypass capacitor between the screen and the cathode. Suppose, then that the grid-cathode voltage is held fixed, and the plate-cathode voltage is increased by a small amount E_{pc}. The tube current then will increase by an amount I_{Epc}. The ratio between these quantities is, then, equal to the first of the tube constants, plate resistance R_p so that:

$$R_p = \frac{E_{pc}}{I_{Epc}}$$

$$I_{Epc}R_p = E_{pc}$$

and,

$$I_{Epc} = \frac{E_{pc}}{R_p}$$

If, on the other hand, the plate cathode voltage is held fixed and the grid-cathode voltage is increased by an amount E_{gc}, the resulting change in current is I_{Egc}. The ratio between these quantities, with the current in the numerator, is the second of the tube constants, the mutual conductance, G_m. Then as a result,

$$G_m = \frac{I_{Egc}}{E_{gc}}$$

and,

$$I_{Egc} = E_{gc}G_m$$

Amplification factor μ is the ratio of E_{pc} to E_{gc} for equal changes in the tube current. Thus:

$$\mu = \frac{E_{pc}}{E_{gc}}$$

$$E_{pc} = \mu E_{gc}$$

if the value of E_{pc} and E_{gc} are chosen so that I_{Epc} and I_{Egc} are equal. Since:

$$I_{Epc} = \frac{E_{pc}}{R_p}$$

and,

$$I_{Egc} = E_{gc}G_m$$

under all circumstances; if:

$$I_{Epc} = I_{Egc}$$

then,

$$\frac{E_{pc}}{R_p} = \frac{E_{gc}}{G_m}$$

and,

$$E_{pc} = E_{gc}R_pG_m$$

From the definition of μ, however, E_{pc} is equal to μE_{gc} if E_{pc} and E_{gc} are given values which make I_{Epc} equal to E_{Egc}. Therefore, μ is equal to R_pG_m.

For small changes in the plate-cathode voltage *and* in the grid-cathode voltage, the resulting change in the tube current is the sum of I_{Epc} and I_{Egc}. If then, I_s is the total change in the tube current, which is the *signal* current, then,

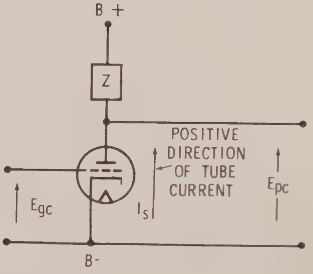
$$I_s = I_{Epc} + I_{Egc}$$

and,

$$I_s = \left(\frac{1}{R_p}\right) E_{pc} + (G_m)\, E_{gc}$$

Now, consider the circuit shown in Fig. II-4, where the arrows show the positive directions for I_s, E_{pc} and E_{gc}. From the circuit, it is clear that E_{pc}, the change in the plate-cathode voltage, must be equal to $-I_sZ$. The equation for the behavior of the tube then becomes:

Fig. II-4. Basic amplifier circuit.



$$I_s = -\left(\frac{Z}{R_p}\right) I_s + (G_m) E_{gc}$$

or,

$$R_pI_s = -ZI_s + R_pG_mE_{gc}$$

or,

$$(R_p + Z)\, I_s = R_pG_m\, E_{gc} = \mu E_{gc}$$

The signal current, then, is expressed by the equation:

$$I_s = \frac{\mu E_{gc}}{(R_p + Z)}$$

Assuming that the B supply is well bypassed, the upper end of load Z is at the same potential, so far as the signal is concerned, as the cathode. The circuit for the tube then may be redrawn as in Fig. II-5, neglecting the DC voltages and currents. As before, the signal current is expressed by the equation:

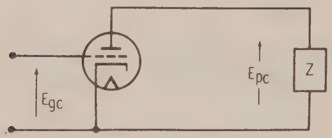


Fig. II-5. Equivalent amplifier circuit.

$$I_s = \frac{\mu E_{gc}}{(R_p + Z)}$$

The tube is, therefore, equivalent to the generator shown in A Fig. II-6. This is a generator having an emf of E_{gc}, which works into a load $(R_p + Z)$. Since the external load consists only of *Z*, the remainder, which is R_p, must be the internal impedance of the generator. Note, however, that a positive current is one that makes the plate end of the lead more negative in relation to the cathode. It is therefore customary to replace this equivalent generator with the one shown in B of Fig. II-6, in which the positive directions of both current and generator voltage are reversed. A positive current then makes the output voltage (E_{pc}) more positive. This equivalent generator is the one generally used as the equivalent of a vacuum tube.

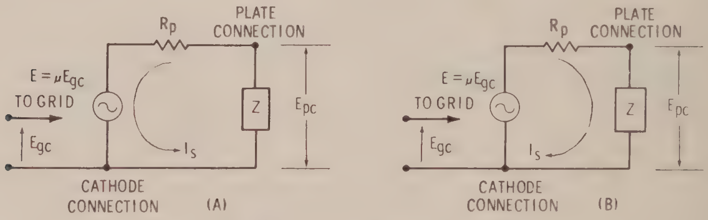

Rp
PLATE CONNECTION
E = μEgc
TO GRID
Egc
Z
Epc
Is
CATHODE CONNECTION
(A)
Rp
PLATE CONNECTION
E = μEgc
TO GRID
Egc
Z
Epc
Is
CATHODE CONNECTION
(B)

Fig. II-6. Equivalent generator circuits.

The equivalent generator shown in Fig. II-6 is used conveniently to replace a triode, since R_p for a triode is usually less than 10,000 ohms (which is something like Z) and μ is reasonably constant, ranging from 20 to 100. A pentode, however, normally has a plate resistance of more than 100,000 ohms—in some cases more than 1 megohm—and a μ which is not always predictable. It is desirable, therefore, to have another equivalent circuit which makes use of G_m in place of μ and in which R_p appears in such a way that minor variations in it will have little effect. Such an equivalent circuit may be found by returning to the equation for the generator shown in Fig. II-5. This is:

$$I_s = -\frac{\mu E_{gc}}{(R_p + Z)}$$

which may be rewritten as:

$$I_s\,(R_p + Z) = -\mu E_{gc} = -R_p G_m E_{gc}$$

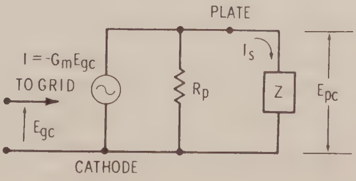
or as,

$$-G_m E_{gc} = I_s\left(\frac{R_p + Z}{R_p}\right) = I_s\left(\frac{1 + Z}{R_p}\right)$$

The right-hand side of this equation is the sum of two terms, I_s and I_sZ/R_p. The first term is, obviuosly, the current through the load. The second term is the current through a resistance, R_p, in parallel with the load. This may be seen by noting that the voltage across the load is I_sZ, whence I_sZ/R_p is the current flowing through a resistor, R_p, having the same voltage across it as does the load. The equation therefore states that the total current through Z and R_p in parallel is equal to $-G_mE_{gc}$.

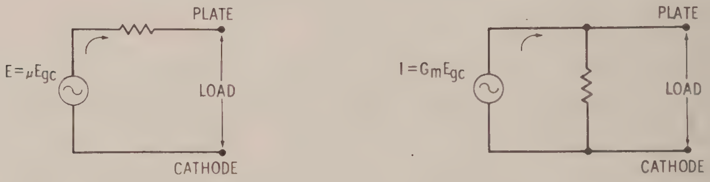
The current equation leads to the equivalent circuit shown in Fig. II-7. Note that this generator delivers current rather than voltage. The voltage generator develops a voltage across its terminals, has a zero impedance, and is connected in series with its internal impedance, R_p. The current generator de-

Fig. II-7. Equivalent current generator circuit.



velops a current *through* itself, has an infinite impedance (so that the current through it is not affected by the voltage across it), and is shunted by its internal impedance R_p. This equivalent generator is very useful for replacing a pentode, since the high R_p of the pentode is usually so much greater than *Z* that it can be neglected, leaving only the current generator and the load, *Z*. In addition, R_p normally draws such a small fraction of the total current that variation in R_p makes little difference in the voltage appearing across *Z*.

A tube having a grid-cathode voltage (E_{gc}) can be replaced, as shown above, by the voltage generator shown in A of Fig. II-8, or by the current generator shown in B of Fig. II-8, where the load is the entire impedance between the plate and the cathode. Note, however, that E_{gc} is the voltage between grid and cathode, *not* the voltage between grid and ground. The two are, of course, identical when the cathode is bypassed at the signal frequency. They are not, however, in such circuits as the cathode follower, nor in any other circuits where there is a voltage difference between cathode and ground.



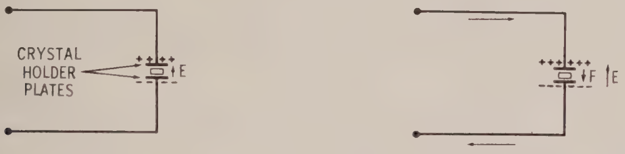
(A) Voltage. (B) Current.

Fig. II-8. Voltage and current generator circuits.

Equivalent Circuit of a Quartz Crystal

A quartz crystal exhibits the phenomenon of *piezoelectricity*. This is simply a technical way of saying that voltage is developed between opposite faces of a quartz crystal when the crystal is deformed mechanically. A number of other substances have the same property, and these are used in crystal pickups for phonographs and in other places where it is necessary to convert mechanical motion into electrical signals. The reverse effect is also present. That is, a voltage across the crystal will cause mechanical distortion.

The quartz crystal used as the frequency-controlling element of a crystal oscillator is a small slab of quartz mounted between a pair of electrodes. The dimensions of the quartz are chosen so that it vibrates mechanically at a frequency very close to the desired oscillator frequency. The general appearance of the crystal and the method of mounting are shown in Fig. II-9A and B.



(A) At rest. (B) Current during deformation.

Fig. II-9. Quartz crystal in mountings.

Before developing the electrical circuit which is equivalent to the quartz crystal, it is desirable to examine the behavior of a single frequency. Consider the pendulum shown in Fig. II-10. This consists of a weight at the end of a light rigid rod. The top of the rod is supported by a hinge system which is reasonably, but not completely, free from friction. Assume that the pendulum has a resonant frequency of *f* cycles per second. If it is started swinging it will oscillate at frequency *f* and the oscillation gradually will die out because of friction at the hinge.

In order to drive the pendulum at its resonant frequency, it is necessary to supply a driving force that will overcome the friction at the supporting hinge. Since the friction always acts in such a direction as will oppose the motion of the pendulum, the driving force must always act in the direction that will assist the motion. In other words, the required driving force is in phase with the pendulum velocity and reverses its direction at the same instant the pendulum velocity is reversed.

The time required for the pendulum to travel from one end of its path to the other is, of course, one half-cycle, and it reaches its midposition (but does not stop) one quarter-cycle after its velocity is reversed at the end of the path. Left to itself, it would take T/4 seconds after velocity reversal to reach the center position, where T is 1/f, and *f* is the resonant frequency. If the pendulum is to be driven at a lower frequency, the time for one quarter-cycle is greater than T/4 second, and it is necessary to hold the pendulum back for some time after its velocity is reversed. The required driving force for a frequency below resonance therefore reverses its direction *after* the pendulum velocity is reversed, so that the driving force and the pendulum velocity are not in phase with each other. The phase of the velocity leads that of the driving force, since the direction of the velocity is reversed earlier than that of the driving force.

Fig. 11-10. Pendulum.

In the extreme case, when the frequency of the driving force is nearly zero, force must be applied, say, to the rght during the entire half-cycle that the pendulum is displaced to the right. The force then reverses as the pendulum passes through the center point of its path, or one quarter-cycle after the velocity is reversed. Thus, the maximum phase difference between the velocity and the driving force, for frequencies below resonance, is 90°.

By a similar reasoning, the velocity lags the driving force for frequencies above resonance. Left to itself, the pendulum comes to rest and reverses its velocity T/4 seconds after passing through the center point of its swings. In order to shorten this time, and reduce the duration of a quarter-cycle the driving force must reverse its direction *before* the pendulum comes to rest so that the pendulum may be slowed up and stopped in less than T/4 second from the time it passes its midpoint.

At extremely high frequencies, the driving force begins to slow that pendulum immediately after the center of the swing is reached, and the maximum phase difference between the velocity and the driving force is again 90°.

Finally, if the maximum velocity of the pendulum is held the same while the driving frequency is varied, the amplitude of the driving force will be a minimum at the resonant frequency. The force is minimized at this frequency because no work is done on the pendulum; the driving force is required only to overcome the friction in the hinge.

The quartz crystal used to control an oscillator may vibrate in a number of different ways at a number of different frequencies. In the neighborhood of any one of the resonant frequencies, however, the crystal may be considered as a system like the pendulum with a single resonant frequency. Since the driving force for oscillation of the crystal is the applied voltage, it is possible to conclude that the velocity of the crystal, as it is deformed mechanically, will lead the applied voltage by angles up to 90° at frequencies below resonance; and will lag behind the applied voltage by angles up to 90° at frequencies above resonance. In addition, for a given velocity at various frequencies, the applied voltage will be a minimum at resonance.

Now, consider what happens when voltage is applied across a crystal. Fig. II-9 shows a crystal between the plates of a crystal holder. Here, the voltage has been applied to the plates, but the crystal has not yet been deformed. The applied voltage is equal to *E* and is in the direction shown by the arrow. As the crystal deforms, it develops a voltage *F*, across itself in a direction that opposes *E*. Clearly, it is necessary to increase the charge on the plates when the crystal is distorted if the voltage between them is to remain the same as when the crystal is undistorted. The current flow to the plates is proportional to the rate at which the charge is increased, and this in turn is proportional to the rate at which the crystal is distorted, or proportional to the crystal velocity.

The required circuit is, therefore, one in which the current leads the voltage below resonance, is in phase with the voltage at resonance, and lags the voltage at frequencies above resonance. In addition, the maximum phase angle between current and voltage is 90°, leading or lagging, and for constant current the voltage is a minimum at resonance. In Fig. II-11, *L*, *R*, and *C* form precisely such a circuit, and *C′* is the capacitance between the plates of the crystal holder.

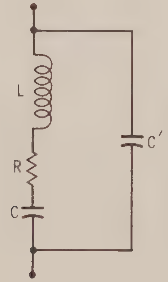


Fig. II-11. Quartz-crystal equivalent circuit.

Index

A-scope, 123
Absorption of energy, 23-24
Absorption wavemeter, 118-121
Acorn tubes, 154
Advantage of waveguide, 83
AFC, 205, 209-212
AGC, 205-207
Airborne radar antenna, 13
Amplification, IF, 197-199
Amplifiers, grounded-grid, 195-197
Analyzer, spectrum, 174
Angle of incidence, 31
Anomalous propagation, 20-23
Antenna
  airborne, 13
  array, 239-255
  early-warning, 13
  microwave, 13, 221-261
Applegate diagram, 165-166
Applications of reflex klystrons, 174
Array method of antenna design, 239-255
Assemblies, frequency-converter, 191
Atmospheric conditions, 19-24
ATR switch, 123, 127-131
Automatic-tracking range gate, 202-203

Bandwidth and directivity of microwave systems, 12-13
Barkhausen oscillator, 144, 161-163
Basic theory of rectangular resonator, 108-109
Beacons, radar, 191-192
Behavior of microwaves, 14-50
Box, echo, 21
B-scope, 124
Buncher, 167

Capacitance, tube, 152-157
Catcher, 168
Cavity magnetrons, 174-185
  operaton
    modes of, 182-185
    theory of, 180-182
Cavity, resonant, 107-108
Cavity resonators, 101-131
  coupling, 117-118
    loop, 118
    probe, 117
    slit, 118
  tuning of, 115-116
Characteristic impedance, 85
Characteristic wavelength of rectangular resonators, 110-111
Characteristics of rectangular waveguide, 79
Characteristics, propagation, 14-18
Circular waveguide, 58
  dimensions, 82
  modes, 79-82
    definition of, 80
Coaxial line, 52-58
  combined fields in, 56-57
  electric field in, 55
  electromagnetic wave in, 57-58
  magnetic field in, 55-56
  S-band, 53
  X- and K-band, 53-55
Colpitts oscillator, 150-151
Combined fields in coaxial line, 56-57
Combined field patterns in rectangular waveguide, 63-66
Communications
  receivers, 212-220
    frequency sharing in, 213-215
    mixed systems, 216-217
    special aspects of, 217-220
      hybrid operation, 219-220
      preselection, 218-219
    systems, point-to-point
      over land, 35-41
      over water, 46
    time sharing in, 215-216
Composition of radio wave, 28
Conditions
  nonstandard, 20
  standard, 20

Conical-scanning system, 200-201
Conventional-tube oscillators, 144
Conversion, frequency, 187
Coupler, directional, 99-100
Coupling
  from mixer to IF amplifier, 198-199
  to cavity resonator, 117-118
    loop, 118
    probe, 117
    slit, 118
Crystal oscillator, 151-152, 153
Curvature of microwave paths, 18-19, 23
Curve, index of refraction, 21
Cut-off wavelength of rectangular waveguide, 66-67
CW radar, 204-205
Cylindrical resonators, 113-115
  field patterns in, 114

Definition of circular-waveguide modes, 80
Density modulation, 164
Detection and amplification, 199-205
Dimensions
  microwave-component, 10-11
  of circular waveguide
  of rectangular resonators
Directional coupler, 99-100
Directivity
  of microwave systems, 12-13
  transmitting-antenna, 26-27
Direct wave, 15-16
Disc-seal tubes, 159-161
Dissipative terminations, 224-225
Double-stub section, 85
Drift tube, 168
Duct, radio-wave, 21

Early-warning antenna, 13
Echo box, 21, 121-127
  grass in, 123
Effect of wave paths, 31-34
Electric field in coaxial line, 55
Electric field patterns in rectangular waveguide, 61-62
Electrode, keep-alive, 130-131
Electromagnetic wave in coaxial line, 57-58
Electron sorting, 162-163
Energy absorption, 23-24
Error in profile diagram, 40-41
Establishing steady oscillation in feedback oscillators, 135-137
Equivalent circuit of single-tube amplifier, 274-178

Feedback oscillators, 133-137
  loop gain of, 134-135
  establishing steady oscillation in, 135-137
Feed systems, reflector, 248-255
Field patterns
  in rectangular waveguide, 61-66
    combined, 63-66
    electric, 61-62
    half-sine, 65
    magnetic, 63
  of rectangular resonator modes, 112-113
Fields, coaxial-line, 55-58
  combined, 56-57
  electric, 55
  magnetic, 55-56
Finding of VSWR, 90
FM radar, 203-204
Frequency
  image, 193
  sharing, 213-215
Frequency conversion, 187
Frequency-converter assemblies, 191
Frequency-multiplier oscillators, 144

Generator-to-line impedance matching, 222
Grass, 123
Grounded-grip amplifiers, 195-197
Group velocity, 70-72
Guided waves, 51
Guide wavelength, 101-102
  of rectangular waveguide, 67-68

Half-sine field pattern in rectangular waveguide, 65
Hartley oscillator, 150
Height-range diagrams, 41-43
High-power, three-reactance oscillators, 155-157
Horizontally-polarized waves, 28-31
  reflection of, 30
Hybrid operation, 218-219

Image
  frequency, 193
  rejection, 192
Impedance
  characteristic, 85, 93-94
  matching
    generator-to-line, 222
    line-to-load, 221-222
    line-to-space, 223
Importance of reflected signal, 24-26
Incidence, angle of, 31
Index of refraction, 19-20
  curve, 21
Inductance, lead, 152-157
Intermediate-frequency amplification, 197-199
Interpretation of radiation patterns, 237-239
Inversion
  radiation, 22
  surface, 22
  temperature, 21

Joint, rotating, 91-93

K-band coaxial line, 53-55
K-band waveguide, 11
Keep-alive electrode, 130-131
Kipp relay, 141-142
Klystron, 132-133
  oscillators, 166-174
  reflex, 171-174
    applications of, 174
    modulation techniques, 172-173
    pulsed operation, 173-174

LC circuit, resonant, 104-105
Lead inductance, 152-157
Lenses, metal, 255-257
Lighthouse tubes, 159-161
Light waves, speed of, 19
Limitations
  line-of-sight, 34-35
  on maximum range, 46-50
Line
  lossy, 224
  stretcher, 91, 129
Line-of-sight limitations, 34-35
Line source power gain, 233-234
Line-to-load impedance matching, 221-223
Line-to-space impedance matching, 223
Linear array power gain, 236-237
Load, sand, 225
Lobe-switching system, 199
Local oscillator, 191
  radiation, 194
Loop coupling to cavity resonator, 118
Loop gain of feedback oscillators, 134-135
Lossy line, 224
Lower sideband, 12

Magnetic field in coaxial line, 55-56
Magnetic field patterns in rectangular waveguide, 63
Magnetron, 132-133
M-curve, 21-22
Maximum operating range, 48-50
Maximum-range limitations, 46-50
Metal lenses, 255-257
Microwave
  antennas, 13, 221-261
    airborne, 13
    early-warning, 13
  aspects, 9-10
  behavior, 14-50
  component dimensions, 10-11
  frequency versus wavelength, 11
  introduction, 7-13
  receivers, 186-220
    frequency conversion in, 187
    local oscillator used in, 191
    mixers used in, 187-191

    - preselection of, 192-197
    - sensitivity of, 193-194
  - signals
    - curvature of, 18-19, 23
  - systems
    - bandwidth and directivity of, 12-13
  - transmission systems, 51-100
- Mixed systems, 216-217
- Mixer-IF amplifier coupling, 198-199
- Mixers, receiver, 187-191
- Modes
  - circular-waveguide, 58
    - of cavity-magnetron operation, 182-185
  - rectangular-resonator, 108-109
    - field patterns of, 112-113
    - terminology of, 111-112
  - rectangular waveguide, 68-79
    - dominant TE, 80
  - terminology of, 69
    - transverse electric (TE), 69
    - transverse magnetic (TM), 69
- Modulation
  - density, 164
  - velocity, 163-166
- Multivibrator relaxation oscillator, 140-141

- One-shot multivibrator, 141-142
- Operating modes of rectangular resonators, 108-109
  - field patterns of, 112-113
  - terminology of, 111-112
- Operating range, maximum, 48-50
- Operating theory of the cavity magnetron, 180-182
- Operation
  - over land, 25-26
  - over water, 25-26
- Oscillators, 132-185
  - Barkhausen, 144, 161-163
  - conventional-tube, 144
  - feedback, 133-137
    - establishing steady oscillation in, 135-137
    - loop gain of, 134-135
  - klystron, 166-174
    - reflex, 171-174
      - applications of, 174
      - modulation techniques, 172-173
      - pulsed operation of, 173-174
  - local, 191
    - radiation of, 194
  - RC, 145-146, 152-153
  - relaxation, 137-142
    - multivibrator circuit, 140-141
    - nonelectronic, 142
    - relay, 142
    - thyratron circuit, 138-140
    - univibrator circuit, 141-142
  - ringing, 142-144
    - frequency-multivibrator, 144
    - tri-tet, 143-144
  - three-reactance, 146-152, 153
    - Colpitts, 150-151
    - crystal, 151-152, 153
    - Hartley, 150-154
    - tuned-plate, tuned-grid, 149-150
- Other resonators, 115
- Overland operation, 25-26
  - of point-to-point communications system, 35-41
  - radar coverage, 43-46
- Overwater operation, 25-26
  - of point-to-point communications system, 46

- Parabolic reflector, 240-242
- Paraboloid of revolution, 242-243
- Paths of radio waves, 15
  - direct, 15
  - reflected, 15
  - sky, 15
  - surface, 15
- Pattern velocity, 70-72
- Phase velocity, 70-72
- Point-to-point communication system
  - over land, 35-41
  - over water, 46
- Polar diagram, vertical, 33-34
- Power
  - delivered to the receiver, 48

  density of at receiving site, 47-48
  gain
    linear array, 236-237
    line source, 233-234
PPI-scope, 124
Preselection, 192-197
Pre-TR switch, 128
Principles of resonant circuits, 104-108
Probe coupling to cavity resonator, 117
Profile diagram, 36-40
  error in, 40-41
Propagation characteristics, 14-18
Pulsed operation of reflex klystron, 173-174
Pulse-type radar, 199-202
  systems used in
    conical-scanning, 200-201
    lobe-switching, 199
Preselection, 219-220
Propagation, anomalous, 20-23
Pythagoras' theorm, 44

Quarter-wave stub supports, 53
Quartz-crystal equivalent circuit, 278-282

Racon, 191-192
Radar, 7
  beacons, 191-192
  coverage, 41-43
    graph of, 43
    over land, 43-44
    over water, 44-46
  CW, 204-205
  FM, 203-204
  maximum operating range, 49-50
  pulse, 199, 202
  search, 27-28
  tracking, 26-27
Radiation
  local oscillator, 194
  patterns, 225-237
    interpretation of, 237-239
Radio wave
  composition of, 28
  duct, 21
  speed of, 19
Range, maximum operating, 48-50
RC oscillator, 145-146
Receivers
  communications, 212-220
    frequency sharing in, 213-215
    mixed systems, 216-217
    special aspects of, 217-220
      hybrid operation, 218-219
      preselection, 219-220
  microwave, 186-220
    frequency conversion in, 187
    local oscillator used in, 191
    mixer used in, 187-191
    preselection of, 192-197
    sensitivity of, 193-194
Rectangular resonators, 108-109
  basic theory of, 108
  characteristic wavelengths of, 110-111
  dimensions of, 110
  field patterns of modes in, 112-113
  operating modes of, 109
  terminology of modes in, 111-112
Rectangular waveguide, 58
  characteristics, 79
  cut-off wavelength, 66-67
  field configuration of complete pattern, 77
  field patterns, 61-66
    combined, 63-66
    electric, 61-62
    half-sine, 65
    magnetic, 63
  guide wavelength, 67-68
  single-mode transmission, 78
  transmission, 58-79
    modes of, 68-79
    dominant TE, 80
Re-entrant cavities, 115
Reflected wave, 15-16
Reflection
  horizontally polarized wave, 30
  vertically polarized wave, 30-31
Reflector
  feed systems, 248-255
  parabolic, 240-242
Reflex klystron, 171-174

  applications of, 174
  modulation techniques, 172-173
  pulsed operation of, 173-174
Refraction, index of, 19-20
  curve, 21
Reinforcement of signals, 101-102
Repection, image, 192
Relation between X, Y and λ_g, 103
Relaxation oscillators, 137-142
  multivibrator circuit, 140-141
  nonelectronic, 142
  relay, 142
  thyratron circuit, 138-140
  univibrator circuit, 141-142
  one-shot multivibrator, 141-142
  Kipp relay, 141-142
Relay relaxation oscillator, 142
Resonant cavity, 107-108
Resonant circuits
  LC, 104-105
  principles of, 104-108
  transmission-line section, 105-107
Resonators
  cavity, 101-131
  cylindrical, 113-115
    field patterns in, 114
  simple, 102-104
Ringing oscillators, 142-144
  tri-tet, 144
  frequency multipliers, 144
Ringing time, 123-124
  test, 126
Rotating joint, 91-93

Sand load, 225
S-band
  coaxial line, 53
  waveguide, 11
Scanning and stabilization, 257-261
Scattering of signal, 24
Search-and-beacon frequency converter assembly, 192
Sensitivity-time-control system, 207-209
Sidebands
  lower, 12
  upper, 12
Signal
  reinforcement, 101-102
  scattering, 24
Simple resonator, 102,104
Single-tube amplifier equivalent circuit, 274-278
Sky wave, 15, 17
Slotted section, 88-90, 98-99
Slug tuner, 94-96
Sorting, electron, 162-163
Special reflectors, 246-248
Spectrum analyzer, 174
Speed
  light waves, 19
  radio waves, 19
Stabilization, 257-261
Standard conditions, 20
Standard waveguides, 10
Stretcher, line, 91, 129
Stub supports, quarter-wave, 53
Stubs, tuning, 84-85, 96-97
  adjustment of, 88
Surface wave, 15, 18
Supports, quarter-wave stub, 53
Switch
  ATR, 123, 127-131
  pre-TR, 128
  TR, 123, 127-131

$TE_{o,n}$ modes in rectangular waveguide, 72-74
Temperature inversion, 21-22
Terminations, dissipative, 224-225
Terminology of rectangular-resonator modes, 111-112
Theory of transmission line, 85-88
Three-reactance oscillators, 146-152
Thyratron relaxation oscillator, 138-140
Time sharing, 215-216
$TM_{o,n}$ mode in rectangular waveguide, 75
TR switch, 123, 127-131
Tracking radar, 26-27
Transformer-coupled oscillator, 145
Transmission

  rectangular waveguide, 58-79
    modes in, 68-79
  systems
    microwave, 51-100
Transmission line
  resonant section, 105-107
  theory, 85-88
  types of, 52-55
Transmitting-antenna directivity, 26-27
Transport velocity, 70-72
Transverse electric (TE) terminology, 69
Transverse magnetic (TM) terminology, 69
Tri-tet oscillator, 143-144
Tube
  acorn, 154
  capacitance of, 152-157
  disc seal, 159-161
  drift, 168
  lighthouse, 159-161
Tunable-cavity wavemeter, 120-121
Tuned-plate, tuned-grid oscillator, 149-150
Tuner, slug, 94-96
Tuning devices and related equipment, 83-100
  directional coupler, 99-100
  double-stub section, 85
  line stretcher, 91
  rotating joint, 91-93
  slotted section, 85-90, 98-99
  slug tuner, 94-96
  tuning stubs, 84-85, 96-97
Tuning plug and paddle, 116
Tuning stubs, 84-85, 96-97
  adjustment of, 88
Two nonexistant modes, 70
Types of cavity resonators, 118-126
Types of transmission lines, 52-55

Unguided waves, 51-52
Univibrator relaxation oscillator, 141-142
Upper sideband, 12

Velocity
  group, 70-72
  modulation, 163-166
  pattern, 70-72
  phase, 70-72
  transport, 70-72
  variaton of with wavelength, 71-72
Vertical polar diagram, 33-34
Vertically-polarized waves, 28-31
  reflection of, 30-31
VSWR (voltage standing wave ratio), 89
  finding of, 90

Wave front, 18-19
Waveguide, 52, 58-79
  advantages, 83
  circular, 58
  cut-off wavelength, 66-67
  guide wavelength, 67-68
  impedance, 93-94
  rectangular, 58
    transmission in, 58-79
      modes of, 68-79
    field patterns in, 61-66
      combined, 63-66
      electric, 61-62
      half-sine, 65
      magnetic, 63
  standard, 10
    K-band, 11
    S-band, 11
    X-band, 11
Wavemeters, 118-121
  absorption, 121
  tunable cavity, 120-121
Wave paths, effect of , 31-34
Waves
  guided, 51
  unguided, 51-52

X-band
  coaxial line, 53-55
  waveguide, 11

MICROWAVE SYSTEMS
FUNDAMENTALS

by F. Jonathan Mivec

For the past 20 years, Mr. Mivec has been a member of International Business Machines Corporation, setting up and servicing all types of electronic computers. In 1959 he was assigned duties in the new field of tele-processing. This assignment, requiring much research in the area of microwaves as used in the transmission of data, made him acutely aware of the lack of an all-inclusive text covering microwave principles and fundamentals. As a result, he formulated the idea of a book on the subject to aid others in the same situation.

Microwave Systems Fundamentals fills the growing need for information on microwave systems, both radar and communications. Beginning with a study of microwave behavior, the text discusses microwave radio paths, the effects of atmospheric conditions on microwaves, direct and reflected signals, overland and overwater communications, line-of-sight limitations, and maximum ranges. Also discussed are guided and unguided waves, types of transmission lines, and waveguides and their mechanical and electrical properties. All the hardware that goes into the making of low-loss systems is fully illustrated and explained.

One chapter is devoted entirely to the study of various types of cavity resonators, their principles and applications. Another chapter, on microwave oscillators, presents klystrons and magnetrons—going into great detail to explain their principles and applications. The chapter on microwave receivers is a complete analysis of frequency conversion, preselection amplifiers, and AGC and AFC. In addition, simplex and duplex systems are discussed, along with frequency and time sharing, and hybrid operation.

Yet another chapter, devoted to nothing but a study of microwave antennas, shows you the how and why of transmission-line and dissipative terminations. The basic fundamentals of the transmission and reception of radiation patterns are thoroughly laid out and illustrated. Likewise, the various types of radiators and reflectors are presented and discussed, along with their various types of feeds.

All in all, you will find no other text on the market that will give you such complete technical information on microwave systems and their uses. Not only will you find *Microwave Systems Fundamentals* easy to understand when you first study it, but you will want to pick it up again and again to brush up on a subject that is growing so rapidly in this day of missiles and satellites.

®

HOWARD W. SAMS & CO., INC.

THE BOBBS-MERRILL COMPANY, INC.

$5.95
MSM-1